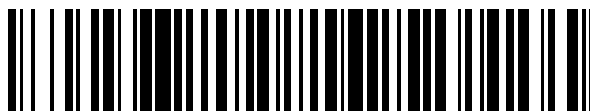


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 625 144**

51 Int. Cl.:

C07K 14/01 (2006.01)

C07H 21/04 (2006.01)

C12N 7/01 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **25.09.2008** **PCT/NZ2008/000248**

87 Fecha y número de publicación internacional: **02.04.2009** **WO09041831**

96 Fecha de presentación y número de la solicitud europea: **25.09.2008** **E 08834643 (2)**

97 Fecha y número de publicación de la concesión europea: **14.12.2016** **EP 2203467**

54 Título: **Polinucleótidos y polipéptidos del fargo MRU y usos de éstos**

30 Prioridad:

25.09.2007 US 975104 P

22.11.2007 US 989840 P

22.11.2007 US 989841 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
18.07.2017

73 Titular/es:

PASTORAL GREENHOUSE GAS RESEARCH LTD
(100.0%)

Level 13, 113-119 The Terrace
Wellington 6011, NZ

72 Inventor/es:

ALTERMANN, ERIC HEINZ;
ATTWOOD, GRAEME TREVOR;
LEAHY, SINEAD CHRISTINE;
KELLY, WILLIAM JOHN;
RONIMUS, ROBERT STARR;
LI, DONG;
KONG, ZHANHAO;
SCHOFIELD, LINLEY ROSE;
DEY, DEBJIT;
TOOTILL, CATHERINE MARY;
SANG, CARRIE;
MOON, CHRISTINA DIANE y
JANSSEN, PETRUS, HENDRICUS

74 Agente/Representante:

ZEA CHECA, Bernabé

ES 2 625 144 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Polinucleótidos y polipéptidos del fago mru y usos de éstos

Campo de la invención

La presente descripción se refiere a composiciones y métodos para administrar moléculas inhibitoras en células microbianas, en particular, células metanógenas. Específicamente, la descripción se refiere a fago ϕ mru identificado recientemente, incluyendo inducción de fago, partículas de fago, y el genoma del fago, y también a polipéptidos del fago, así como polinucleótidos que codifican estos polipéptidos. La descripción también se refiere a vectores de expresión y células huésped para producir estos polipéptidos. La descripción se refiere además a métodos para detectar, tomar como diana, e inhibir células microbianas, especialmente células metanógenas, usando el fago, polipéptidos, polinucleótidos, vectores de expresión, y células huésped descritos.

Antecedentes de la invención

En Nueva Zelanda, la actividad agrícola es responsable de la mayoría de las emisiones de gases de efecto invernadero. Por lo tanto, la reducción de las emisiones agrícolas de gases de efecto invernadero es importante para cumplir con las obligaciones de Nueva Zelanda bajo el Protocolo de Kioto. El Protocolo requiere la reducción de gases de efecto invernadero a niveles de 1990 hacia el final del primer periodo de compromiso (2008-2012). Para este fin, los grupos del sector agrícola y el gobierno de Nueva Zelanda establecieron el Pastoral Greenhouse Gas Research Consortium (PGGRC) para identificar medios para reducir las emisiones de gases de efecto invernadero agrícolas en Nueva Zelanda.

Una parte importante de las actividades del PGGRC ha sido la investigación en la reducción de las emisiones de metano por los rumiantes de pastoreo de Nueva Zelanda. El mitigar las emisiones de metano de rumiantes tiene un interés comercial por dos razones. En primer lugar, el fracaso en cumplir con los compromisos del Protocolo de Kioto forzará al gobierno a adquirir créditos de carbón. Esto se estima que cuesta actualmente 350 millones de dólares. En segundo lugar, la producción de metano resulta en la pérdida de 8-12% de la energía bruta producida en el rumen. Esta energía podría usarse, en su lugar, para mejorar la productividad de los rumiantes.

El metano se produce en el rumen por microbios denominados metanógenos que son parte del filo *Euryarchaeota* en el reino *Archaea*. La mayor parte de los metanógenos crecen en CO₂ y H₂ como su única fuente de energía, pero algunos pueden usar acetato o compuestos metilo para crecer. Varios diferentes géneros de arquea metanogénicos existen en el rumen, pero se piensa que las especies del género *Methanobrevibacter*, especialmente *M. ruminantium*, y *M. smithii* son los metanógenos predominantes en los rumiantes de Nueva Zelanda. *M. ruminantium* es actualmente el objeto de un proyecto de secuenciación de genoma financiado por el PGGRC. El proyecto es la primera secuenciación de genoma de un metanógeno del rumen y tiene como objetivo obtener una mejor comprensión de la biología de *Methanobrevibacter* para descubrir dianas para la inhibición de la formación de metano.

La reducción de la producción de metano en el rumen requiere la inhibición de metanógenos o la inactivación de su ruta de metanogénesis. Wright et al., Vaccine, vol. 22, no. 29-30, páginas 3976-3985, 2004 describe un método para inhibir una célula metanogénica por inmunización de un animal frente a metanógenos del rumen. Un medio para inhibir la producción de metano es administrar moléculas inhibitoras específicas en las células metanógenas. Esto puede conseguirse, por ejemplo, por el uso de agentes, tales como bacteriófagos, que tienen como diana específicamente los metanógenos. Se han caracterizado varios fagos para metanógenos no de rumen, pero no ha habido informes publicados de fagos capaces de infectar o lisar metanógenos del rumen. Por lo tanto, sería altamente ventajoso identificar fagos que tengan la capacidad de infectar células metanógenas y/o administrar inhibidores.

Resumen de la invención

Aquí, se describe un fago ϕ mru aislado, incluyendo una partícula de fago y/o genoma de fago, producido completamente o en parte, así como polinucleótidos y polipéptidos aislados del fago como se describe con detalle en la presente memoria.

Aquí, también se describe un polipéptido aislado que comprende al menos una secuencia de aminoácidos de fago seleccionada del grupo que consiste en SEQ ID NO: 1-69. En particular se describe, el polipéptido comprende la secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 2-5 y 62-68. Se describe además, el polipéptido comprende la secuencia de aminoácidos de SEQ ID NO: 63. Aquí se describe, el polipéptido es un fragmento, por ejemplo, que comprende al menos una secuencia de aminoácidos que se extiende desde los residuos 32-186 de SEQ ID NO: 63.

En un aspecto, la invención presenta un polipéptido aislado: a) que comprende SEQ ID NO: 72; b) que es una variante funcional de SEQ ID NO: 72, y que comparte al menos 90% de identidad con la secuencia de longitud completa de SEQ ID NO: 72; o c) que es una variante funcional de SEQ ID NO: 72, y que comparte al menos 80%

de identidad con la secuencia de longitud completa de SEQ ID NO: 72; en el que dichas variantes funcionales son capaces de inhibir una célula metanógena.

La invención describe adicionalmente un polinucleótido aislado que comprende una secuencia codificadora para al menos un polipéptido del fago. Se describe, el polinucleótido comprende una secuencia codificadora para al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69. Se describe en particular, el polinucleótido comprende una secuencia codificadora para una secuencia seleccionada del grupo que consiste en SEQ ID NO: 2-5 y 62-68. Se describe además, el polinucleótido comprende una secuencia codificadora para SEQ ID NO: 63. Aquí se describe, el polinucleótido comprende un fragmento de una secuencia codificadora, por ejemplo, al menos una secuencia de aminoácidos que se extiende de los residuos 32-186 de SEQ ID NO: 63.

En un aspecto, la invención presenta adicionalmente un polinucleótido aislado que comprende: a) secuencia codificadora para la secuencia de longitud completa de SEQ ID NO: 72; b) secuencia codificadora para un polipéptido que comparte al menos 90% de identidad con la secuencia de longitud completa de SEQ ID NO: 72, y que es una variante funcional de ésta; c) secuencia codificadora para un polipéptido que comparte al menos 80% de identidad con la secuencia de longitud completa de SEQ ID NO: 72, y que es una variante funcional de ésta; o d) una secuencia complementaria a una cualquiera de (a) a (c), en el que dichas variantes funcionales son capaces de inhibir una célula metanógena.

Aquí, se describe un polinucleótido aislado que comprende una secuencia de ácido nucleico de fago seleccionada del grupo que consiste en SEQ ID NO: 74-142. En particular, el polinucleótido comprende una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 75-78 y 135-141, o es particularmente, SEQ ID NO: 136. Aquí, se describe el polinucleótido es un fragmento o un oligonucleótido que comprende, por ejemplo, la secuencia de ácido nucleico que se extiende desde los nucleótidos 94-558 de SEQ ID NO: 136. Además, aquí se describe un polinucleótido aislado, o fragmento de éste, que hibrida con una cualquiera de las secuencias de ácido nucleico de SEQ ID NO: 74-142. Aquí, se describe además un polinucleótido aislado que comprende el complemento, complemento reverso, secuencia reversa, o fragmentos de éstos, de una cualquiera de las secuencias de ácido nucleico.

Aquí, se describe el vector de expresión comprende una secuencia codificadora para al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69. En particular, se describe el vector de expresión comprende una secuencia codificadora para al menos una secuencia de aminoácidos de SEQ ID NO: 2-5 y 62-68. Se describe además, el vector de expresión comprende una secuencia codificadora para al menos una secuencia de aminoácidos de SEQ ID NO: 63. Se describe además, el vector de expresión comprende una secuencia codificadora para al menos una secuencia de aminoácidos que se extiende desde los residuos 32-186 de SEQ ID NO: 63.

En un aspecto, la invención se refiere a un vector a) que codifica el polipéptido de la invención; o b) que comprende el polinucleótido de la reivindicación 1 o reivindicación 2.

La invención también presenta una célula huésped, por ejemplo, una célula huésped microbiana, que comprende al menos un vector de expresión que codifica el polipéptido de la reivindicación 1 o que comprende el polinucleótido de la reivindicación 2. Aquí se describe un anticuerpo producido frente a un polipéptido o polinucleótido como se describe en la presente memoria. Aquí también se describe, el anticuerpo está dirigido a al menos una secuencia de polipéptido seleccionada del grupo que consiste en SEQ ID NO: 1-69, o una secuencia modificada de ésta. Aquí también se describe, el anticuerpo se produce frente al menos un fragmento de un polinucleótido seleccionado del grupo que consiste en SEQ ID NO: 74-142, o un complemento, o secuencia modificada de ésta. En otro aspecto, el anticuerpo incluye una o más fusiones o conjugados con al menos un inhibidor celular, por ejemplo, compuestos anti-metanogénesis (por ejemplo, ácido bromoetanosulfónico), anticuerpos y fragmentos de anticuerpo, enzimas líticas, ácidos nucleicos peptídicos, péptidos antimicrobianos, y otros antibióticos como se describe con detalle en la presente memoria.

Aquí se describen polipéptidos de fago modificados, por ejemplo, para al menos una de SEQ ID NO: 1-69, incluyendo alteraciones, fragmentos, variantes, y derivados biológicamente activos, descritos en la presente memoria. Aquí se describen adicionalmente anticuerpos modificados, por ejemplo, dirigidos a al menos una de SEQ ID NO: 1-69, incluyendo alteraciones, fragmentos, variantes, y derivados biológicamente activos, descritos en la presente memoria. También se presentan polinucleótidos que codifican estos polipéptidos modificados, así como alteraciones, fragmentos, variantes, y derivados de los polinucleótidos descritos, vectores de expresión que comprenden estos polinucleótidos, y células huésped que comprenden estos vectores. En aspectos específicos, las composiciones y métodos de la invención emplean estos polinucleótidos o polipéptidos modificados, o vectores de expresión o células huésped correspondientes.

Además, aquí se describen polipéptidos de fago, por ejemplo, al menos una de SEQ ID NO: 1-69 o secuencias modificadas de éstas, que incluyen fusiones o conjugados con al menos un inhibidor celular, por ejemplo, compuestos anti-metanogénesis (por ejemplo, ácido bromoetanosulfónico), anticuerpos y fragmentos de anticuerpo, enzimas líticas, ácidos nucleicos peptídicos, péptidos antimicrobianos, y otros antibióticos como se describe con detalle en la presente memoria. En un aspecto, la invención se refiere a una molécula conjugada o molécula de

fusión que comprende el polipéptido de la invención, y que comprende además un compuesto anti-metanogénesis, una secuencia señal, un anticuerpo o fragmento de anticuerpo, un ácido nucleico peptídico, un péptido antimicrobiano, o un antibiótico.

5 Aquí también se describe una composición que comprende un polipéptido aislado, por ejemplo, al menos una de SEQ ID NO: 1-69, o una secuencia modificada de ésta. Aquí se describe una composición que comprende un anticuerpo, por ejemplo, dirigido frente al menos una de SEQ ID NO: 1-69, o una secuencia modificada de ésta. También se presenta una composición que comprende un polinucleótido aislado, por ejemplo, al menos una de SEQ ID NO: 74-142, o un complemento o secuencia modificada de ésta. Además, se presenta una composición que incluye un vector de expresión, o célula huésped que comprende un vector de expresión, según la invención. La
10 composición puede incluir una cualquiera de las alteraciones, fragmentos, variantes, y derivados biológicamente activos descritos en la presente memoria. Las composiciones pueden incluir al menos un inhibidor celular (por ejemplo, como una fusión o conjugado), y pueden formularse, por ejemplo, como composiciones farmacéuticas o como suplementos alimenticios, en particular, componentes alimenticios de rumiantes.

15 Aquí también se describe una composición como parte de un kit para tomar como diana y/o inhibir células microbianas, especialmente células metanógenas, según los métodos descritos. Los kits comprenden: a) al menos una composición como se muestra en la presente memoria; y b) opcionalmente, instrucciones para uso, por ejemplo, para tomar como diana células o inhibir el crecimiento celular o replicación para metanógenos u otros microbios.

20 Aquí también se describe un método para producir un fago, comprendiendo el método: a) cultivar un vector de expresión o célula huésped que comprende un vector de expresión, que comprende al menos parte del genoma del fago en condiciones adecuadas para la producción del fago; y b) recuperar el fago del cultivo. En aspectos particulares, el fago comprende al menos un polipéptido seleccionado del grupo que consiste en SEQ ID NO: 1-69, o secuencias modificadas de éstas. Se describe además, el fago comprende al menos un polinucleótido seleccionado del grupo que consiste en SEQ ID NO: 74-142, o secuencias modificadas de éstas.

25 Aquí también se describe un método para producir un polipéptido de fago, comprendiendo el método: a) cultivar un vector de expresión o célula huésped que comprende un vector de expresión, que comprende al menos parte de una secuencia codificadora para al menos un polipéptido de fago en condiciones adecuadas para la expresión del polipéptido; y b) recuperar el polipéptido del cultivo. En particular se describe, el polipéptido comprende al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, o secuencias modificadas de éstas.

30 Aquí se describe adicionalmente un método para producir un polipéptido de fago, por ejemplo, para al menos una de SEQ ID NO: 1-69, que comprende una fusión o conjugado con al menos un inhibidor celular, por ejemplo, compuestos anti-metanogénesis (por ejemplo, ácido bromoetanosulfónico), anticuerpos y fragmentos de anticuerpo, enzimas líticas, ácidos nucleicos peptídicos, péptidos antimicrobianos, y otros antibióticos como se describe con detalle en la presente memoria. Dicho método comprende: a) cultivar un vector de expresión o célula huésped que
35 comprende un vector de expresión, que comprende una secuencia codificadora para al menos un polipéptido de fago en condiciones adecuadas para la expresión del polipéptido; b) formar la fusión o conjugado de fago (por ejemplo, por expresión de la secuencia fusionada o conjugación química con el inhibidor celular); y c) recuperar la fusión o conjugado. En particular, el polipéptido comprende al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, o secuencias modificadas de éstas.

40 La invención presenta un método para inhibir una célula metanógena que comprende poner en contacto la célula con el polipéptido de la reivindicación 1, poner en contacto la célula con la molécula conjugada o molécula de fusión de la reivindicación 7 u 8, o con el fago de la reivindicación 9. Además, la invención presenta un método para inhibir (por ejemplo, inhibir el crecimiento o replicación) una célula microbiana, en particular, una célula metanógena, que comprende: a) opcionalmente, producir o aislar al menos un polipéptido de fago; y b) poner en contacto la célula con el polipéptido de fago. Aquí también se describe, el polipéptido comprende al menos una secuencia de aminoácidos
45 seleccionada del grupo que consiste en SEQ ID NO: 1-69, o una secuencia modificada de éstas.

Como una característica añadida, la presente descripción engloba un método para inhibir (por ejemplo, inhibir el crecimiento o replicación) una célula microbiana, en particular, una célula metanógena, que comprende: a) opcionalmente, producir o aislar al menos un polipéptido de fago, que comprende además al menos un inhibidor celular; y b) poner en contacto la célula con el polipéptido de fago. Aquí también se describe, el polipéptido
50 comprende al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, o una secuencia modificada de éstas.

55 Aquí también se describe un método para detectar y/o medir los niveles de un fago, o un polipéptido o polinucleótido de fago correspondiente, que comprende: 1) poner en contacto una muestra de un sujeto con un anticuerpo producido frente a un polipéptido de fago (por ejemplo, al menos una de SEQ ID NO: 1-69, o una secuencia modificada de éstas) o un polinucleótido correspondiente; y 2) determinar la presencia o niveles del complejo de anticuerpo formado con el polipéptido o polinucleótido en la muestra. Dichos métodos también pueden usarse para detectar y/o medir los niveles de una célula microbiana, en particular, una célula metanógena.

Aquí también se describe un método para detectar y/o medir los niveles de un fago, o un polinucleótido de fago correspondiente (por ejemplo, una secuencia codificadora de fago), que comprende: 1) poner en contacto una muestra de un sujeto con un polinucleótido complementario (por ejemplo, una secuencia complementaria a una cualquiera de SEQ ID NO: 74-142, o secuencia modificada de éstas); y 2) determinar la presencia o niveles del complejo de hibridación formado con el polinucleótido de fago en la muestra. Dichos métodos también pueden usarse para detectar y/o medir los niveles de una célula microbiana, en particular, una célula metanógena.

En aspectos particulares, los métodos de la invención utilizan componentes de expresión *in vivo* o *in vitro*. En otros aspectos, los métodos emplean polipéptidos producidos por medios recombinantes, sintéticos, o semi-sintéticos, o polipéptidos producidos por medios endógenos.

Otros aspectos y realizaciones de la invención se describen en la presente memoria a continuación.

Descripción breve de los dibujos

Esta invención se describe con referencia a realizaciones específicas de ésta y con referencia a las figuras.

FIGS. 1A-1B. Profago ϕ mru de *M. ruminantium* que muestra las secuencias del sitio de integración posibles *attL* y *attR* (FIG. 1A), y los módulos funcionales y estructura génica predichos del fago (FIG. 1B).

Fig 1C: Inducción de fago usando aire estéril (estrés de oxígeno).

Fig 1D: Inducción de fago inicial usando MitomicinaC.

Fig 1 E: Electroforesis en gel de agarosa de amplicones de PCR de *M. ruminantium* inducido (pulso de oxígeno) y no inducido. Carriles 2 y 4: escalera de ADN de marcador de 1 kb por Invitrogen. Carriles 1 y 3 representan PCR usando la pareja de cebadores R1F-L2R en ADN aislado de cultivos de *M. ruminantium* inducido y no inducido, respectivamente.

FIG. 2. Anotaciones y comentarios del marco de lectura abierto del profago ϕ mru.

FIG. 3. Anotación, función predicha y comentarios del marco de lectura abierto del profago ϕ mru.

FIGS. 4A-4B. Información de secuencia del profago ϕ mru de *M. ruminantium*, incluyendo secuencias codificadoras del fago ϕ mru (FIG. 4A), y secuencias de aminoácidos del fago ϕ mru (FIG. 4B).

FIG. 5. Alineamiento de secuencias de ORF 2058 del fago ϕ mru con PeiP de *M. marburgensis* y PeiW de *M. wolfeii*.

FIG. 6: Logo de secuencia de proteína de secuencias de péptido señal de *M. ruminantium* creado usando LogoBar, que muestra la señal consenso central.

FIG. 7: Efecto inhibitor de ORF 2058 en células de *M. ruminantium* en reposo.

FIG. 8: Efecto inhibitor de ORF 2058 en el crecimiento celular y producción de metano de *M. ruminantium*.

Descripción detallada de la invención

Definiciones

Secuencias de ácidos nucleicos "alteradas" que codifican polipéptidos de fago, tal y como se usa en la presente memoria, incluyen aquellas con deleciones, inserciones, o sustituciones de diferentes nucleótidos que resultan en un polinucleótido que codifica preferiblemente los mismos polipéptidos o funcionalmente equivalentes. El polipéptido o anticuerpo codificado también puede estar "alterado" y contener deleciones, inserciones, o sustituciones de residuos de aminoácidos que producen un cambio silencioso y resultan en un polipéptido funcionalmente equivalente. Las sustituciones de aminoácidos deliberadas pueden hacerse sobre la base de similitud en polaridad, carga, solubilidad, hidrofobicidad, hidrofiliidad, y/o la naturaleza anfipática de los residuos siempre que la actividad biológica (por ejemplo, asociación celular, permeabilización celular, o lisis celular) o actividad inmunológica (por ejemplo, uno o más sitios de unión de anticuerpo) del polipéptido se retenga. Por ejemplo, los aminoácidos cargados negativamente pueden incluir ácido aspártico y ácido glutámico; los aminoácidos cargados positivamente pueden incluir lisina y arginina; y los aminoácidos con grupos de cabeza polares no cargados que tienen valores de hidrofiliidad similares pueden incluir leucina, isoleucina, y valina, glicina y alanina, asparagina y glutamina, serina y treonina, y fenilalanina y tirosina.

"Secuencia de aminoácidos", tal y como se usa en la presente memoria, se refiere a una secuencia de oligopéptido, péptido, polipéptido, o proteína, y a cualquier fragmento de éstas, y a moléculas naturales, recombinantes, sintéticas, o semi-sintéticas. Las secuencias de la descripción comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250 aminoácidos, preferiblemente al menos 5 a 10, 10 a 20, 20 a 30, 30 a 40, 40 a 50, 50 a 100, 100 a 150, 150 a 200, ó 200 a 250, ó 250 a 4.000 aminoácidos, y, preferiblemente, retienen la actividad biológica (por ejemplo, asociación celular, permeabilización celular, o lisis celular) o actividad inmunológica (por ejemplo, uno o más sitios de unión de anticuerpo) de la secuencia original. Donde la "secuencia de aminoácidos" se

recita en la presente memoria para hacer referencia a una secuencia de aminoácidos de una molécula de polipéptido natural, la secuencia de aminoácidos, y términos semejantes, no se pretende que limite la secuencia de aminoácidos a la secuencia de aminoácidos completa original asociada con la molécula de longitud completa.

- 5 "Amplificación", tal y como se usa en la presente memoria, se refiere a la producción de copias adicionales de una secuencia de ácido nucleico y se lleva a cabo generalmente usando tecnologías de la reacción en cadena de la polimerasa (PCR) muy conocidas en la técnica (Dieffenbach, C. W. y G. S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

- 10 El término "anticuerpo" debe entenderse en el sentido más amplio posible y se pretende que incluya anticuerpos monoclonales y anticuerpos policlonales intactos. También se pretende que abarque fragmentos y derivados de anticuerpos siempre que presenten la actividad biológica deseada. Anticuerpos engloba moléculas de inmunoglobulina y partes inmunológicamente activas de moléculas de inmunoglobulina (Ig), es decir, moléculas que contienen un sitio de unión a antígeno que se une específicamente (inmunorreacciona con) un antígeno. Éstos incluyen, pero no están limitados a, policlonales, monoclonales, quiméricos, de cadena única, fragmentos Fc, Fab, Fab', y Fab₂, y una biblioteca de expresión de Fab.

- 15 Las moléculas de anticuerpo se refieren a cualquiera de las clases IgG, IgM, IgA, IgE, e IgD, que se diferencian entre sí por la naturaleza de la cadena pesada presente en la molécula. Éstas incluyen también subclases, tales como IgG1, IgG2, y otras. La cadena ligera puede ser una cadena kappa o una cadena lambda. La referencia en la presente memoria a anticuerpos incluye una referencia a todas las clases, subclases, y tipos. También están incluidos los anticuerpos quiméricos, por ejemplo, anticuerpos monoclonales o fragmentos de éstos que son
20 específicos para más de una fuente, por ejemplo, una o más secuencias de ratón, humanas, o de rumiante. Están incluidos además anticuerpos de camélidos o nanocuerpos. Se entenderá que cada referencia a "anticuerpos" o cualquier término semejante, en la presente memoria incluye anticuerpos intactos, así como cualesquiera fragmentos, alteraciones, derivados, o variantes de éstos.

- 25 Los términos "biólogicamente activo" o "funcional", tal y como se usan en la presente memoria, se refieren a un polipéptido que retiene una o más funciones estructurales, inmunológicas, o bioquímicas (por ejemplo, asociación celular, permeabilización celular, o lisis celular) de la secuencia.

- Los términos "inhibidor celular" o "inhibidor," tal y como se usan en la presente memoria, se refieren a agentes que disminuyen o bloquean el crecimiento o replicación de células microbianas, especialmente células metanógenas. Un inhibidor celular puede actuar para disminuir o bloquear, por ejemplo, la división celular. Un inhibidor puede disminuir o bloquear, por ejemplo, la síntesis de ADN, síntesis de ARN, síntesis de proteínas, o modificaciones posteriores a la traducción. Un inhibidor también puede disminuir o bloquear la actividad de enzimas implicadas en la ruta de la metanogénesis. Un inhibidor también puede estar dirigido a una célula para el reconocimiento por los componentes del sistema inmune. La inhibición de una célula también incluye el matar a la célula y la muerte celular, por ejemplo, de lisis, apoptosis, necrosis, etc. Los inhibidores útiles incluyen, pero no están limitados a, compuestos anti-metanogénesis (por ejemplo, ácido bromoetanosulfónico), anticuerpos y fragmentos de anticuerpo, enzimas líticas,
30 ácidos nucleicos peptídicos, péptidos antimicrobianos, y otros antibióticos como se describe con detalle en la presente memoria.

- Los términos "complementario" o "complementariedad," tal y como se usan en la presente memoria, se refieren a la unión natural de polinucleótidos bajo condiciones permisivas de sal y temperatura por emparejamiento de bases. Para la secuencia A-G-T, la secuencia complementaria es T-C-A, el complemento reverso es A-C-T y la secuencia reversa es T-G-A. La complementariedad entre dos moléculas monocatenarias puede ser parcial, en la que sólo algunos de los ácidos nucleicos se unen, o puede ser completa cuando existe la complementariedad total entre las moléculas monocatenarias. El grado de complementariedad entre cadenas de ácido nucleico tiene efectos significativos en la eficiencia y fuerza de la hibridación entre cadenas de ácido nucleico. Esto es particularmente importante en las reacciones de amplificación, que dependen de la unión entre cadenas de ácido nucleico y en el diseño y uso de moléculas de PNA.
40 45

- El término "derivado", tal y como se usa en la presente memoria, se refiere a la modificación química de un ácido nucleico que codifica un polipéptido de fago, o un ácido nucleico complementario a éste. Dichas modificaciones incluyen, por ejemplo, el reemplazo de hidrógeno por un grupo alquilo, acilo o amino. En aspectos preferidos, un derivado de ácido nucleico codifica un polipéptido que retiene la función biológica o inmunológica de la molécula natural. Un polipéptido derivado es uno que está modificado por glicosilación, pegilación, o cualquier proceso similar que retiene una o más funciones biológicas (por ejemplo, asociación celular, permeabilización celular, o lisis celular) o función inmunológica de la secuencia de la que se derivó.
50

- El término "homología", tal y como se usa en la presente memoria, se refiere a un grado de complementariedad. Puede haber homología parcial (es decir, I de identidad) u homología completa (es decir, 100% de identidad). Una secuencia parcialmente complementaria que inhibe al menos parcialmente la hibridación de una secuencia idéntica con un ácido nucleico diana se refiere usando el término funcional "sustancialmente homóloga". La inhibición de la hibridación de la secuencia completamente complementaria a la secuencia diana puede examinarse usando un ensayo de hibridación (transferencia Southern o Northern, hibridación en disolución y semejantes) en condiciones de
55

baja astringencia. Una secuencia sustancialmente homóloga o sonda de hibridación competirá para e inhibirá la unión de una secuencia completamente homóloga a la secuencia diana en condiciones de baja astringencia. Esto no significa que las condiciones de baja astringencia son tales que se permita la unión no específica; las condiciones de baja astringencia requieren que la unión de dos secuencias entre sí sea una interacción específica (es decir, selectiva).

El término "hibridación", tal y como se usa en la presente memoria, se refiere a cualquier proceso mediante el que una cadena de ácido nucleico se une con una cadena complementaria a través de emparejamiento de bases.

Una "inserción" o "adición", tal y como se usa en la presente memoria, se refiere a un cambio en una secuencia de aminoácidos o nucleótidos que resulta en la adición de uno o más residuos de aminoácidos o nucleótidos, respectivamente, comparado con la molécula natural.

Un "metanógeno," tal y como se usa en la presente memoria, se refiere a microbios que producen gas metano, que incluyen *Methanobrevibacter*, *Methanothermobacter*, *Methanomicrobium*, *Methanobacterium*, y *Methanosarcina*. Los metanógenos específicos incluyen, pero no están limitados a, *Methanobrevibacter ruminantium*, *Methanobrevibacter smithii*, *Methanobrevibacter acididurans*, *Methanobrevibacter thaueri*, *Methanobacterium bryantii*, *Methanobacterium formicicum*, *Methanothermobacter marburgensis*, *Methanothermobacter wolfeii*, *Methanosphaera stadtmanae*, *Methanomicrobium mobile*, *Methanosarcina barkeri*, *Methanosarcina mazei*, *Methanococcoides burtonii*, y *Methanobolus taylorii*. Todos los géneros y especies de metanógenos están englobados por este término.

Células "microbianas" tal y como se usa en la presente memoria, se refiere a células microbianas naturales o modificadas genéticamente incluyendo arqueobacterias tales como metanógenos, halófilos, y termoacidófilos, y eubacterias, tales como cianobacteria, espiroqueta, proteobacteria, así como bacterias gram positivas y gram negativas.

El término "modificado" se refiere a secuencias alteradas y a fragmentos, variantes, y derivados de secuencias, como se describen en la presente memoria.

"Secuencia de ácido nucleico" o "secuencia de nucleótidos" tal y como se usa en la presente memoria, se refiere a una secuencia de un polinucleótido, oligonucleótido, o fragmentos de éstos, y a ADN o ARN de origen natural, recombinante, sintético, o semi-sintético que puede ser mono o bicatenario, y puede representar la cadena con sentido o antisentido, y regiones codificadoras o no codificadoras. Las secuencias de la descripción incluyen lo más preferiblemente secuencias codificadoras de polipéptido que comprenden al menos 12, 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 300, 450, 600, 750 nucleótidos, preferiblemente al menos 15 a 30, 30 a 60, 60 a 90, 90 a 120, 120 a 150, 150 a 300, 300 a 450, 450 a 600, ó 600 a 750 nucleótidos, o al menos 1.000 nucleótidos, o al menos 1.500 nucleótidos. Se entenderá que cada referencia a una "secuencia de ácido nucleico" o "secuencia de nucleótidos" en la presente memoria, incluirá la secuencia de longitud completa original, así como cualesquiera complementos, fragmentos, alteraciones, derivados, o variantes, de éstos.

El término "oligonucleótido" se refiere a una secuencia de ácido nucleico que comprende al menos 6, 8, 10, 12, 15, 18, 21, 25, 27, 30, ó 36 nucleótidos, o al menos 12 a 36 nucleótidos, o al menos 15 a 30 nucleótidos, que puede usarse, por ejemplo, en amplificación por PCR, secuenciación, o ensayos de hibridación. Tal y como se usa en la presente memoria, oligonucleótido es sustancialmente equivalente a los términos "amplímeros", "cebadores", "oligómeros", "oligos", y "sondas", como se define comúnmente en la técnica.

"Polipéptido," tal y como se usa en la presente memoria, se refiere a los polipéptidos aislados de la invención obtenidos de cualquier especie, preferiblemente microbianos, de cualquier fuente ya sea natural, sintética, semi-sintética, o recombinante. Específicamente, un polipéptido de fago puede obtenerse de células metanógenas, tales como células de *Methanobrevibacter*, en particular, células de *M. ruminantium*, o *M. smithii*. Para la producción recombinante, un polipéptido de la invención puede obtenerse de células microbianas o eucariotas, por ejemplo, células de *Escherichia*, *Streptomyces*, *Bacillus*, *Salmonella*, levadura, insecto tales como *Drosophila*, células animales tales como células COS y CHO, o células de plantas. Se entenderá que cada referencia a un "polipéptido", en la presente memoria, incluirá la secuencia de longitud completa original, así como cualesquiera fragmentos, alteraciones, derivados, o variantes, de éstas.

El término "polinucleótido", cuando se usa en el singular o plural, generalmente se refiere a cualquier secuencia de ácido nucleico, por ejemplo, cualquier polirribonucleótido o polidesoxirribonucleótido, que puede ser ARN o ADN no modificado o ARN o ADN modificado. Esto incluye, sin limitación, ADN mono y bicatenario, ADN que incluye regiones mono y bicatenarias, ARN mono y bicatenario, y ARN que incluye regiones mono y bicatenarias, moléculas híbridas que comprenden ADN y ARN que pueden ser mono catenarias o, más típicamente, bicatenarias o incluyen regiones mono y bicatenarias. También se incluyen regiones tricatenarias que comprenden ARN o ADN o tanto ARN como ADN. Específicamente incluidos están ARNm, ADNc, y ADN genómicos, y cualesquiera fragmentos de éstos. El término incluye ADN y ARN que contienen una o más bases modificadas, tales como bases tritiadas, o bases inusuales, tales como inosina. Los polinucleótidos de la invención pueden englobar secuencias codificadoras o no codificadoras, o secuencias con sentido o antisentido, o ARNi tales como ARNsi. Se entenderá que cada referencia a un "polinucleótido" o término semejante, en la presente memoria, incluirá las secuencias de longitud completa, así como cualesquiera complementos, fragmentos, alteraciones, derivados, o variantes de éstos.

"Ácido nucleico peptídico" o "PNA" tal y como se usa en la presente memoria, se refiere a una molécula antisentido o agente anti- gen que comprende bases unidas a través de un núcleo peptídico.

El término "rumiante", tal y como se usa en la presente memoria, se refiere a animales que tienen un rumen como un tipo especial de órgano digestivo. Los rumiantes incluyen, pero no están limitados a, ganado, ovejas, cabras, búfalo, alce, antílope, caribú, y ciervo.

Los términos "condiciones astringentes" o "astringencia", tal y como se usan en la presente memoria, se refieren a las condiciones para hibridación definidas por el ácido nucleico, sal, y temperatura. Estas condiciones son muy conocidas en la técnica y pueden alterarse con el fin de identificar o detectar secuencias de polinucleótido idénticas o relacionadas. Véase, por ejemplo, Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, NY, y Ausubel, F. M. et al. (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, Nueva York, NY. Numerosas condiciones equivalentes que comprenden bien baja o alta astringencia dependen de factores tales como la longitud y naturaleza de la secuencia (ADN, ARN, composición de bases), naturaleza de la diana (ADN, ARN, composición de bases), medio (en disolución o inmovilizado en un sustrato sólido), concentración de sales y otros componentes (por ejemplo, formamida, sulfato de dextrano y/o polietilenglicol), y temperatura de las reacciones (en un intervalo de aproximadamente 5°C por debajo de la temperatura de fusión de la sonda a aproximadamente 20°C a 25°C por debajo de la temperatura de fusión). Uno o más factores pueden variarse para generar condiciones bien de baja o alta astringencia diferentes de, pero equivalentes a, las condiciones listadas anteriormente.

El término "sujeto" incluye animales humanos y no humanos. Los animales no humanos incluyen, pero no están limitados a, pájaros y mamíferos, tales como rumiantes, y en particular, ratones, conejos, gatos, perros, cerdos, ovejas, cabras, vacas, y caballos.

Los términos "sustancialmente purificado" o "aislado" tal y como se usan en la presente memoria, se refieren a secuencias de ácido nucleico o aminoácidos que se retiran de su entorno celular, recombinante, o sintético, y son al menos 60% libres, preferiblemente 75% libres, y lo más preferiblemente al menos 90% libres o al menos 99% libres de otros componentes con los que están asociados en un entorno celular, recombinante, o sintético.

"Transformación", tal y como se define en la presente memoria, describe un proceso mediante el que el ADN exógeno entra en y cambia una célula receptora. Puede ocurrir en condiciones naturales o artificiales usando varios métodos muy conocidos en la técnica. La transformación puede basarse en cualquier método conocido para la inserción de secuencias de ácido nucleico extrañas en una célula huésped procariota o eucariota. El método se selecciona sobre la base del tipo de célula huésped que se está transformando y puede incluir, pero no está limitado a, infección viral, electroporación, choque de calor, lipofección, y bombardeo de partículas. Dichas células "transformadas" incluyen células transformadas de forma estable en las que el ADN insertado es capaz de replicación bien como un plásmido que se replica autónomamente o como parte del cromosoma del huésped. También incluyen células que expresan de forma temporal el ADN o ARN insertado durante periodos de tiempo limitados.

Una "variante" de un polipéptido, tal y como se usa en la presente memoria, se refiere a una secuencia de aminoácidos que se altera por uno o más aminoácidos. Un polinucleótido variante se altera por uno o más nucleótidos. Una variante puede resultar en cambios "conservativos", en los que un aminoácido sustituido tiene propiedades estructurales o químicas similares, por ejemplo, reemplazo de leucina con isoleucina. Más raramente, una variante puede resultar en cambios "no conservativos", por ejemplo, reemplazo de una glicina con un triptófano. Las variaciones menores análogas también pueden incluir delecciones o inserciones de aminoácidos, o ambas. La guía para determinar qué residuos de aminoácidos pueden sustituirse, insertarse, o deleccionarse sin suprimir la actividad biológica o inmunológica puede encontrarse usando programas informáticos muy conocidos en la técnica, por ejemplo, software LASERGENE (DNASTAR).

La descripción también engloba variantes que retienen al menos una actividad biológica (por ejemplo, asociación celular, permeabilización celular, o lisis celular) o actividad inmunológica del polipéptido. Una variante preferida es una que tiene sustancialmente la misma secuencia o una funcionalmente equivalente, por ejemplo, al menos 80%, y más preferiblemente al menos 90% de identidad de secuencia respecto a una secuencia descrita. Una variante lo más preferida es una que tiene al menos 95%, al menos 97%, al menos 98%, al menos 99%, al menos 99,5%, al menos 99,8%, o al menos 99,9% de identidad de secuencia respecto a una secuencia descrita en la presente memoria. El porcentaje de identidad se determina mediante el alineamiento de las dos secuencias que se van a comparar como se describe más adelante, determinando el número de residuos idénticos en la parte alineada, dividiendo ese número por el número total de residuos en la secuencia inventiva (de búsqueda), y multiplicando el resultado por 100. Un programa de alineamiento útil es AlignX (Vector NTI).

Descripción de la invención

El metano se produce en el intestino anterior de los rumiantes por metanógenos que actúan como reductores terminales del carbono en el sistema del rumen. La ruta de metanogénesis de múltiples etapas está bien elucidada, principalmente a partir del estudio de metanógenos no del rumen, pero las adaptaciones que permiten a los metanógenos crecer y persistir en el rumen no se entienden bien. *Methanobrevibacter ruminantium* es un

metanógeno prominente en los rumiantes de Nueva Zelanda. Como se describe en la presente memoria, el genoma de *M. ruminantium* se ha secuenciado y se muestra con un tamaño de aproximadamente 3,0 Mb con un contenido GC de 33,68%. Como un descubrimiento inesperado, se encontró que el genoma de *M. ruminantium* incluye una secuencia de profago (designado ϕ mr_u) con módulos funcionales distintos que codifica la integración del fago, replicación de y encapsidación del ADN, proteínas de la cápside, lisis, y funciones de conversión lisogénicas.

El fago de *M. ruminantium* se identificó durante secuenciación de alto rendimiento, cuando se encontró que una región de 30 a 40 kb del genoma estaba sobre-representada en los clones secuenciados. Esto sugirió que una parte del genoma estaba presente en un número más alto de copias de lo normal, y podría atribuirse a la replicación de un fago residente. La región sobre-representada se investigó y los análisis bioinformáticos detallados de los marcos de lectura abiertos predichos presentes indicaron que contenía genes semejantes a fago. Se ha mostrado que una región con bajo GC encontrada en el extremo distal de la secuencia de fago (conversión lisogénica) alberga un sistema de modificación de ADN predicho por azufre (dnd) que podría proporcionar una modificación adicional de ADN de huésped o extraño. La secuencia de profago de *M. ruminantium* se describe con detalle en la presente memoria. En varios aspectos de la invención, los polinucleótidos y polipéptidos del profago pueden usarse como un medio para inhibir los metanógenos y/o la metanogénesis en el rumen, y para elucidar adicionalmente el papel de *M. ruminantium* en la formación de metano.

Aquí se describen por lo tanto polipéptidos de fago, incluyendo aquellos que comprenden al menos una de SEQ ID NO: 1-69, y fragmentos, variantes, y derivados de éstos. Aquí también se describe el uso de estos polipéptidos para tomar como diana e inhibir células microbianas, especialmente células metanógenas. Aquí se describe además engloba el uso de los polipéptidos para la inhibición del crecimiento o replicación de dichas células. Los polipéptidos de la presente invención pueden expresarse y usarse en varios ensayos para determinar su actividad biológica. Los polipéptidos pueden usarse para protocolos de síntesis y aislamiento a gran escala, por ejemplo, para producción comercial. Dichos polipéptidos pueden usarse para producir anticuerpos, para aislar secuencias de aminoácidos correspondientes, y para determinar cuantitativamente los niveles de las secuencias de aminoácidos. Los polipéptidos de la presente invención también pueden usarse como composiciones, por ejemplo, composiciones farmacéuticas, y como suplementos alimenticios, por ejemplo, componentes del alimento de rumiantes. Los polipéptidos de la presente invención también tienen beneficios para la salud. En aspectos relacionados con la salud, los inhibidores de metanógenos pueden usarse para restaurar la energía en el sujeto que normalmente se pierde como metano. En aspectos particulares, pueden usarse dispositivos ruminales de liberación lenta conjuntamente con los polipéptidos, y composiciones (por ejemplo, composiciones farmacéuticas y suplementos alimenticios) de la descripción.

Aquí también se describen polipéptidos que comprenden al menos una secuencia seleccionada del grupo que consiste en: (a) polipéptidos que comprenden al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, o fragmentos, variantes, o derivados de éstas; (b) polipéptidos que comprenden un dominio funcional de al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, y fragmentos y variantes de éstas; y (c) polipéptidos que comprenden al menos un número especificado de residuos contiguos de al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, o variantes o derivados de éstas.

Aquí también se describen polinucleótidos que codifican al menos un polipéptido de fago, incluyendo aquellos de SEQ ID NO: 1-69, y fragmentos, variantes, y derivados de éstos. Aquí también se describe el uso de estos polinucleótidos para preparar vectores de expresión y células huésped para tomar como diana e inhibir células microbianas, especialmente células metanógenas. Aquí se describe el uso de los polinucleótidos para la inhibición del crecimiento o replicación de dichas células. Los polinucleótidos aislados de la presente invención también tienen utilidad en el mapeo genómico, en el mapeo físico, y en la clonación de genes de fagos más o menos relacionados. Las sondas diseñadas usando los polinucleótidos de la presente invención pueden usarse para detectar la presencia y examinar los patrones de expresión de genes en cualquier organismo que tenga secuencias de ADN y ARN lo suficientemente homólogas en sus células, usando técnicas que son muy conocidas en la técnica, tales como técnicas transferencia en ranura o análisis de micromatrices. Los cebadores diseñados usando los polinucleótidos de la presente invención pueden usarse para secuenciación y amplificaciones por PCR. Los polinucleótidos de la presente invención también pueden usarse como composiciones, por ejemplo, composiciones farmacéuticas, y como suplementos alimenticios, por ejemplo, componentes de alimento de rumiantes. Los polinucleótidos de la presente invención también tienen beneficios para la salud. Para dichos beneficios, los polinucleótidos pueden presentarse como vectores de expresión o células huésped que comprenden vectores de expresión. En aspectos particulares, pueden usarse dispositivos ruminales de liberación lenta conjuntamente con los polinucleótidos, vectores, células huésped, y composiciones (por ejemplo, composiciones farmacéuticas y suplementos alimenticios) de la descripción.

Aquí también se describen polinucleótidos que comprenden al menos una secuencia seleccionada del grupo que consiste en: (a) secuencias que comprenden una secuencia codificadora para al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, o fragmentos o variantes de éstas; (b) complementos, secuencias reversas, y complementos reversos de una secuencia codificadora para al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, o fragmentos o variantes de éstas; (c) marcos de lectura abiertos contenidos en la secuencia codificadora para al menos una secuencia de

aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, y sus fragmentos y variantes; (d) dominios funcionales de una secuencia codificadora para al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, y fragmentos y variantes de éstas; y (e) secuencias que comprenden al menos un número especificado de residuos contiguos de una secuencia codificadora para al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69, o variantes de éstas. Aquí también se describe un polinucleótido aislado que comprende una secuencia codificadora para al menos una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 1-69.

Aquí también se describen polinucleótidos que comprenden al menos una secuencia seleccionada del grupo que consiste en: (a) secuencias que comprenden al menos una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 74-142, o fragmentos o variantes de éstas; (b) complementos, secuencias reversas, y complementos reversos de una secuencia codificadora para al menos una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 74-142, o fragmentos o variantes de éstas; (c) marcos de lectura abiertos contenidos en la secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 74-142, y sus fragmentos y variantes; (d) dominios funcionales de una secuencia codificadora de al menos una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 74-142, y fragmentos y variantes de éstas; y (e) secuencias que comprenden al menos un número especificado de residuos contiguos de al menos una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 74-142, o variantes de éstas. También se proporcionan sondas y cebadores de oligonucleótido y sus variantes obtenidas de cualquiera de las secuencias descritas.

Los expertos en la técnica apreciarán que, como resultado de la degeneración del código genético, puede producirse una multitud de secuencias de nucleótidos que codifican los polipéptidos de la invención, algunas presentando una homología mínima con las secuencias de nucleótidos de cualquier gen conocido y natural. Así, la descripción contempla cada una y todas las variaciones posibles de secuencia de nucleótidos que podrían prepararse mediante la selección de combinaciones sobre la base de elecciones de codones posibles. Estas combinaciones se hacen según el código genético de triplete bacteriano estándar como se aplica a secuencias de aminoácidos naturales.

Las secuencias de nucleótidos que codifican los polipéptidos de fago, o sus fragmentos o variantes, son preferiblemente capaces de hibridar con la secuencia de nucleótidos de la secuencia natural en condiciones apropiadamente seleccionadas de astringencia. Sin embargo, puede ser ventajoso producir secuencias de nucleótidos que codifican un polipéptido, o su fragmento o derivado, que poseen un uso de codones sustancialmente diferente. Pueden seleccionarse codones para incrementar la velocidad a la que ocurre la expresión del polipéptido en un huésped procariota o eucariota particular según la frecuencia con la que los codones particulares se utilizan por el huésped. Por ejemplo, los codones pueden optimizarse para expresión en *E. coli* según métodos conocidos. Otras razones para alterar sustancialmente la secuencia de nucleótidos que codifica polipéptidos y sus derivados sin alterar las secuencias de aminoácidos codificadas incluyen la producción de transcritos de ARN que tienen propiedades más deseables, tales como mayor vida media, que los transcritos producidos a partir de la secuencia natural.

Aquí también se describe la producción de secuencias de ADN, o fragmentos de éstas, que codifican los polipéptidos, o sus fragmentos o variantes, completamente por química sintética. Después de la producción, la secuencia sintética puede insertarse en cualquiera de los muchos vectores de expresión y sistemas de células disponibles usando reactivos que son muy conocidos en la técnica. Además, la química sintética puede usarse para introducir mutaciones en una secuencia que codifica un polipéptido, o cualesquiera variantes o fragmento de ésta. Aquí también se describen secuencias de polinucleótido que son capaces de hibridar con las secuencias de nucleótidos reivindicadas, y en particular, aquellas mostradas en SEQ ID NO: 74-149, o sus complementos en varias condiciones de astringencia como se enseña en Wahl, G. M. y S. L. Berger (1987; *Métodos Enzymol.* 152:399-407) y Kimmel, A. R. (1987; *Métodos Enzymol.* 152:507-511).

Los métodos para la secuenciación de ADN son muy conocidos y están disponibles generalmente en la técnica y pueden usarse para poner en práctica cualquiera de las realizaciones de la descripción. Los métodos pueden emplear enzimas tales como el fragmento de Klenow de la ADN polimerasa I, SEQUENASA (U.S. Biochemical Corp, Cleveland, OH), polimerasa Taq (Perkin Elmer), polimerasa termoestable T7 (Amersham Pharmacia Biotech, Piscataway, NJ), o combinaciones de polimerasas y exonucleasas con capacidad de corrección tales como las encontradas en el sistema de amplificación ELONGASE comercializado por Life Technologies (Gaithersburg, MD). Preferiblemente, el proceso se automatiza con máquinas tales como el Hamilton Micro Lab 2200 (Hamilton, Reno, NV), ciclador termal Peltier (PTC200; MJ Research, Watertown, MA) el catalizador ABI y secuenciadores de ADN 373 y 377 (Perkin Elmer), o el secuenciador Genome 20™ (Roche Diagnostics).

Las secuencias de ácido nucleico que codifican los polipéptidos pueden extenderse utilizando una secuencia de nucleótidos parcial y empleando varios métodos conocidos en la técnica para detectar secuencias en 5' tales como promotores y elementos reguladores, y elementos en 3' tales como terminadores y estructuras de ARN no codificadoras. Por ejemplo, un método que puede emplearse, PCR de "sitio de restricción" usa cebadores universales para recuperar la secuencia desconocida adyacente a un locus conocido (Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322). En particular, el ADN genómico se amplifica en primer lugar en presencia de cebador a una secuencia conectora y un cebador específico de la región conocida. Las secuencias amplificadas se someten

entonces a una segunda ronda de PCR con el mismo cebador conector y otro cebador específico interno al primero. Los productos de cada ronda de PCR se transcriben con una ARN polimerasa apropiada y se secuencian usando transcriptasa inversa.

Otro método útil es PCR, inversa, también denominada IPCR (véase, por ejemplo, Ochman H, Gerber AS, Hartl DL. Genetics. 1988 nov;120(3):621-3). La PCR inversa puede emplearse cuando sólo se conoce una secuencia interna del ADN diana. El método de PCR inversa incluye una serie de digestiones y auto-ligaciones con el ADN siendo cortado por una endonucleasa de restricción. Este corte resulta en una secuencia conocida en uno de los extremos de secuencias desconocidas. Según este método, el ADN diana se corta ligeramente en fragmentos más pequeños de varias kilobases por digestión con endonucleasa de restricción. La auto-ligación se induce entonces bajo concentraciones bajas causando que se reforme el núcleo fosfato y se produzca un producto de ligación de ADN circular. El ADN diana se digiere por restricción con una endonucleasa conocida. Esto genera un corte en la secuencia interna conocida generando un producto lineal con secuencias terminales conocidas. Este producto puede usarse entonces para PCR estándar que se lleva a cabo con cebadores complementarios a las secuencias internas conocidas.

Los sistemas de electroforesis capilar que están disponibles comercialmente pueden usarse para analizar el tamaño o confirmar la secuencia de nucleótidos de los productos de secuenciación o PCR. En particular, la secuenciación capilar puede emplear polímeros en forma de polvo fluido seco para la separación electroforética, cuatro colorantes fluorescentes diferentes (uno para cada nucleótido) que se activan con láser, y detección de las longitudes de onda emitidas por una cámara con dispositivo de carga acoplada. La intensidad de salida/luz puede convertirse en señal eléctrica usando un software apropiado (por ejemplo, GENOTYPER y Sequence NAVIGATOR, Perkin Elmer) y el proceso completo desde la carga de las muestras hasta el análisis informático y presentación de los datos electrónicos puede controlarse por ordenador. La electroforesis capilar es especialmente preferible para la secuenciación de partes pequeñas de ADN que podrían estar presentes en cantidades limitadas en una muestra particular.

Recientemente, la pirosecuenciación ha surgido como una metodología de secuenciación útil. Véase, por ejemplo, Ronaghi, M. et al. 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. 242: 84-89; Ronaghi, M. et al. 1998. A sequencing method based on real-time pyrophosphate. Science 281: 363-365; Ronaghi, M. et al. 1999. Analyses of secondary structures in DNA by pyrosequencing. Anal. Biochem. 267: 65- 71; Ronaghi 2001. Genome Res. Vol. 11, Número 1, 3-11; Nyren The history of pyrosequencing. Métodos Mol Biol. 2007;373:1-14. La pirosecuenciación tiene las ventajas de exactitud, flexibilidad, procesamiento paralelo, y puede automatizarse fácilmente. Además, la técnica dispensa con la necesidad de cebadores marcados, nucleótidos marcados, y electroforesis en gel. Según este método, la polimerasa cataliza la incorporación de nucleótidos en una cadena de ácido nucleico. Como resultado de la incorporación, se liberan moléculas de pirofosfato y posteriormente se convierten por sulfurilasa en ATP. Se produce luz en la reacción de la luciferasa durante la cual se oxida una molécula de luciferina. Después de la adición de cada nucleótido, se lleva a cabo una etapa de lavado para permitir la adición iterativa. Los nucleótidos se degradan continuamente por enzima que degrada nucleótidos permitiendo la adición de un nucleótido posterior. La pirosecuenciación se ha aplicado con éxito como una plataforma para la secuenciación a gran escala, incluyendo análisis genómico y metagenómico (véase, por ejemplo, El secuenciador genómico FLX™ de 454 Life Sciences/Roche).

El sistema SOLiD™ también se ha desarrollado para secuenciación (véase, por ejemplo, Applied Biosystems. Application Fact Sheet para el sistema SOLiD™. Foster City, CA). Esta metodología se basa en la ligación secuencial de oligonucleótidos marcados con colorante para amplificar de forma clonal fragmentos de ADN unidos a lechos magnéticos. En este método, la secuencia de ADN se genera midiendo la ligación seriada. La reacción de ligación se basa en el reconocimiento de sonda, no adición secuencial, y por lo tanto es menos propensa a acumulación de errores. La naturaleza de la química elimina virtualmente la posibilidad de inserciones o deleciones espurias. La etapa de ligación y tratamiento con fosfatasa de las sondas no ligadas previene el desfase. Además, después de siete ciclos de ligación, el cebador original se retira del molde y un nuevo cebador se hibrida para empezar a interrogar en la posición n-1. El uso de esta fase de "reseteado" permite la reducción del ruido sistémico y permite longitudes de lectura más largas. Además, se usa dos bases que codifican para discriminar entre errores de medida en lugar de polimorfismos verdaderos. Los cambios en una única posición se identifican como errores aleatorios y pueden eliminarse por el software en el análisis de los datos. Como una plataforma analítica, el sistema SOLiD™ tiene aplicaciones en la secuenciación a gran escala, expresión génica digital, ChIP y estudios de metilación, y es particularmente útil para detectar variación genómica.

En otra realización de la invención, pueden usarse polinucleótidos o fragmentos de éstos que codifican polipéptidos en moléculas de ADN recombinantes para dirigir la expresión de los polipéptidos, o fragmentos o variantes de éstos, en células huésped apropiadas. Debido a la degeneración inherente del código genético, otras secuencias de ADN que codifican sustancialmente la misma o funcionalmente equivalente secuencia de aminoácidos pueden producirse, y estas secuencias pueden usarse para clonar y expresar polipéptidos de fago. Las secuencias de nucleótidos de la presente invención pueden prepararse por ingeniería usando métodos generalmente conocidos en la técnica con el fin de alterar secuencias que codifican aminoácidos por una variedad de razones, incluyendo, pero no limitado a, alteraciones que modifican la clonación, procesamiento, y/o expresión del producto génico. intercambio de ADN por fragmentación aleatoria y re-ensamblaje por PCR de fragmentos génicos y oligonucleótidos sintéticos puede usarse

para preparar por ingeniería las secuencias de nucleótidos. Por ejemplo, la mutagénesis dirigida a sitio puede usarse para insertar nuevos sitios de restricción, alterar los patrones de glicosilación, cambiar la preferencia de codones, introducir mutaciones, y así sucesivamente.

En otra realización de la invención, las secuencias de ácido nucleico naturales, modificadas, o recombinantes que codifican polipéptidos pueden ligarse a una secuencia heteróloga para codificar una proteína de fusión. Por ejemplo, puede ser útil codificar una secuencia quimérica que pueda ser reconocida por un anticuerpo disponible comercialmente. Una proteína de fusión también puede prepararse por ingeniería para contener un sitio de escisión localizado entre el polipéptido de la invención y la secuencia de la proteína heteróloga, de manera que el polipéptido pueda escindirse y purificarse del resto heterólogo.

En otra realización, las secuencias que codifican polipéptidos pueden sintetizarse, totalmente o en parte, usando métodos químicos muy conocidos en la técnica (véase Caruthers, M. H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232). Alternativamente, el polipéptido en sí mismo puede producirse usando métodos químicos para sintetizar la secuencia de aminoácidos, o un fragmento de ésta. Por ejemplo, la síntesis del polipéptido puede llevarse a cabo usando varias técnicas en fase sólida (Roberge, J. Y. et al. (1995) Science 269:202-204; Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154) y la síntesis automatizada puede conseguirse, por ejemplo, usando el sintetizador peptídico ABI 431A (Perkin Elmer). Varios fragmentos de polipéptidos pueden sintetizarse químicamente separadamente y combinarse usando métodos químicos para producir la molécula de longitud completa.

El polipéptido recién sintetizado puede aislarse por cromatografía líquida de alta resolución preparativa (por ejemplo, Creighton, T. (1983) Proteins Structures y Molecular Principles, WH Freeman and Co., Nueva York, NY). La composición de los polipéptidos sintéticos puede confirmarse por análisis de aminoácidos o secuenciación (por ejemplo, el procedimiento de degradación de Edman; Creighton, supra). Adicionalmente, la secuencia de aminoácidos del polipéptido, o cualquier parte de ésta, puede alterarse durante la síntesis directa y/o combinarse usando métodos químicos con secuencias de otras proteínas, o cualquier parte de éstas, para producir una molécula variante.

Con el fin de expresar polipéptidos biológicamente activos, las secuencias de nucleótidos que codifican el polipéptido o equivalentes funcionales, pueden insertarse en un vector de expresión apropiado, es decir, un vector que contiene los elementos necesarios para la transcripción y traducción de la secuencia codificadora insertada. Los métodos que son muy conocidos para los expertos en la técnica pueden usarse para construir vectores de expresión que contienen secuencias que codifican el polipéptido y elementos de control de la transcripción y traducción apropiados. Estos métodos incluyen técnicas de ADN recombinante *in vitro*, técnicas sintéticas, y recombinación genética *in vivo*. Dichas técnicas se describen en Sambrook, J. et al. (2001) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, y Ausubel, F. M. et al. (2007) Current Protocols in Molecular Biology, John Wiley & Sons, Nueva York, NY.

Puede utilizarse una variedad de sistemas de vector de expresión/huésped para contener y expresar secuencias que codifican los polipéptidos de la invención. Éstos incluyen, pero no están limitados a, microorganismos tales como bacterias transformadas con vectores de expresión recombinantes de ADN de fago, plásmido o cósmido; levaduras transformadas con vectores de expresión de levadura; sistemas de células de insectos infectadas con vectores de expresión de virus (por ejemplo, baculovirus); sistemas de células de plantas transformadas con vectores de expresión de virus (por ejemplo, virus del mosaico de la coliflor, CaMV; virus del mosaico del tabaco, TMV) o con vectores de expresión bacterianos (por ejemplo, plásmidos Ti o pBR322); o sistemas de células de animales. Para las bacterias, los plásmidos útiles incluyen los plásmidos pET, pRSET, pTrcHis2, y pBAD de Invitrogen, los plásmidos pET y pCDF de Novagen, y los plásmidos Director™ de Sigma-Aldrich. Para los metanógenos, los plásmidos útiles incluyen, pero no están limitados a pME2001, pMV15, y pMP1. En particular, puede usarse *Escherichia coli* con el vector de expresión pET. La invención no está limitada por el vector de expresión o célula huésped empelada.

Los "elementos de control" o "secuencias reguladoras" son aquellas regiones no traducidas del vector-potenciadores, promotores, regiones no traducidas en 5' y 3' - que interaccionan con las proteínas de la célula huésped para llevar a cabo la transcripción y traducción. Dichos elementos pueden variar en su fuerza y especificidad. Dependiendo del sistema de vector y huésped utilizado, puede usarse cualquier número de elementos de la transcripción y traducción adecuados, incluyendo promotores constitutivos e inducibles. Por ejemplo, cuando se clona en sistemas bacterianos, pueden usarse promotores inducibles tales como el promotor híbrido lacZ del fagémido BLUESCRIPT (Stratagene, LaJolla, CA) o plásmido pSPORT1 (Life Technologies) y semejantes. El promotor de polihedrina de baculovirus puede usarse en células de insecto. Los promotores o potenciadores derivados de los genomas de células de plantas (por ejemplo, de choque térmico, RUBISCO, y genes de proteínas de almacenamiento) o de virus de plantas (por ejemplo, promotores virales o secuencias líder) pueden clonarse en el vector.

En sistemas bacterianos, pueden seleccionarse varios vectores de expresión dependiendo del uso pretendido para el polipéptido. Por ejemplo, cuando se necesitan grandes cantidades de polipéptido, pueden usarse vectores que dirigen un alto nivel de expresión de proteínas de fusión que se purifican fácilmente. Dichos vectores incluyen, pero

no están limitados a, los vectores de clonación y expresión multifuncionales de *E. coli* tales como BLUESCRIPT (Stratagene), en el que la secuencia que codifica un polipéptido puede ligarse en el vector en marco con secuencias para la Met amino-terminal y los 7 residuos posteriores de β -galactosidasa de manera que se produce una proteína híbrida; vectores pIN (Van Heeke, G. y S. M. Schuster (1989) J. Biol. Chem. 264:5503-5509); y semejantes.

Los vectores pGEX (Promega, Madison, WI) también pueden usarse para expresar los polipéptidos como proteínas de fusión con glutatión-S-transferasa (GST). En general, dichas proteínas de fusión son solubles y pueden purificarse fácilmente de células lisadas por adsorción a lechos de glutatión-agarosa seguido de elución en presencia de glutatión libre. Las proteínas preparadas en dichos sistemas pueden diseñarse para incluir heparina, trombina, o sitios de escisión de la proteasa factor Xa de manera que el polipéptido clonado de interés puede liberarse del resto GST según se desee. En la levadura, *Saccharomyces cerevisiae*, pueden usarse varios vectores que contienen promotores constitutivos o inducibles tales como factor alfa, alcohol oxidasa, y PGH. Para revisiones, véase Ausubel et al. (supra) y Grant et al. (1987) Methods Enzymol. 153:516-544.

También pueden usarse señales específicas de inicio para conseguir una traducción más eficiente de las secuencias que codifican los polipéptidos de la invención. Dichas señales incluyen el codón de inicio ATG y secuencias adyacentes. En los casos en los que las secuencias que codifican un polipéptido, su codón de inicio, y secuencias en 5' se insertan en el vector de expresión apropiado, pueden no necesitarse señales de control de la transcripción o traducción adicionales. Sin embargo, en los casos en los que se inserta sólo la secuencia codificadora, o un fragmento de ésta, deben proporcionarse señales de control de la traducción exógenas incluyendo el codón de inicio ATG. Además, el codón de inicio debe estar en el marco de lectura correcto para asegurar la traducción del inserto completo. Los elementos de traducción y codones de inicio exógenos pueden ser de varios orígenes, tanto natural como sintético. La eficiencia de la expresión puede potenciarse por la inclusión de potenciadores que sean apropiados para el sistema celular particular que se usa, tales como los descritos en la bibliografía (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Además, puede elegirse una cepa de célula huésped por su capacidad de modular la expresión de las secuencias insertadas o para procesar el polipéptido expresado de la manera deseada. Dichas modificaciones de la secuencia incluyen, pero no están limitadas a, acetilación, carboxilación, glicosilación, fosforilación, lipidación, y acilación. El procesamiento posterior a la traducción que escinde una forma "prepro" del polipéptido también puede usarse para facilitar la inserción, plegamiento, y/o función correcta. Diferentes células huésped que tienen una maquinaria celular específica y mecanismos característicos para las actividades posteriores a la traducción están disponibles en la American Type Culture Collection (ATCC; Bethesda, MD) y pueden elegirse para asegurar la modificación y procesamiento correcto de la secuencia. Las células huésped específicas incluyen, pero no están limitadas a, células metanógenas, tales como células de *Methanobrevibacter*, en particular, células de *M. ruminantium*, o *M. smithii*. Las células huésped de interés incluyen, por ejemplo, *Rhodotorula*, *Aureobasidium*, *Saccharomyces*, *Sporobolomyces*, *Pseudomonas*, *Erwinia* y *Flavobacterium*; u otros organismos tales como *Escherichia*, *Lactobacillus*, *Bacillus*, *Streptomyces*, y semejantes. Las células huésped específicas incluyen *Escherichia coli*, que es particularmente idónea para uso con la presente invención, *Saccharomyces cerevisiae*, *Bacillus thuringiensis*, *Bacillus subtilis*, *Streptomyces lividans*, y semejantes.

Hay varias técnicas para introducir ácidos nucleicos en células eucariotas cultivadas *in vitro*. Éstas incluyen métodos químicos (Feigner et al., Proc. Natl. Acad. Sci., USA, 84:7413-7417 (1987); Bothwell et al., Methods for Cloning and Analysis of Eukaryotic Genes, Eds., Jones y Bartlett Publishers Inc., Boston, Mass. (1990), Ausubel et al., Short Protocols in Molecular Biology, John Wiley and Sons, Nueva York, NY (1992); y Farhood, Annal. NY Acad. Sci., 716:23-34 (1994)), el uso de protoplastos (Bothwell, supra) o pulsos eléctricos (Vatteroni et al., Mutn. Res., 291:163-169 (1993); Sabelnikov, Prog. Biophys. Mol. Biol., 62:119-152 (1994); Bothwell et al., supra; y Ausubel et al., supra), uso de virus atenuados (Davis et al., J. Virol. 1996, 70(6), 3781-3787; Brinster et al. J. Gen. Virol. 2002, 83(Pt 2), 369-381; Moss, Dev. Biol. Stan., 82:55-63 (1994); y Bothwell et al., supra), así como métodos físicos (Fynan et al., supra; Johnston et al., Meth. Cell Biol., 43(Pt A):353-365 (1994); Bothwell et al., supra; y Ausubel et al., supra).

La administración exitosa de ácidos nucleicos a tejidos animales puede conseguirse por liposomas catiónicos (Watanabe et al., Mol. Reprod. Dev., 38:268-274 (1994)), inyección directa de ADN o ARN desnudo en tejido muscular animal (Robinson et al., Vacc., 11:957-960 (1993); Hoffman et al., Vacc. 12:1529-1533; (1994); Xiang et al., Virol., 199:132-140 (1994); Webster et al., Vacc., 12:1495-1498 (1994); Davis et al., Vacc. 12:1503-1509 (1994); Davis et al., Hum. Molec. Gen., 2:1847-1851 (1993); Dalemans et al. Ann NY Acad. Sci. 1995, 772, 255-256. Conry, et al. Cancer Res. 1995, 55(7), 1397-1400), y embriones (Naito et al., Mol. Reprod. Dev., 39:153-161 (1994); y Burdon et al., Mol. Reprod. Dev., 33:436-442 (1992)), inyección intramuscular de vacunas de ARN auto replicante (Davis et al., J. Virol. 1996, 70(6), 3781-3787; Balasuriya et al. Vaccine 2002, 20(11-12), 1609-1617) o inyección intradérmica de ADN usando la tecnología de "pistola de genes" (Johnston et al., supra).

Una variedad de protocolos para detectar y medir la expresión de los polipéptidos de la invención, usando bien anticuerpos policlonales o monoclonales específicos para la proteína, son conocidos en la técnica. Los ejemplos incluyen ensayo inmunoabsorbente ligado a enzima (ELISA), radioinmunoensayo (RIA), y separación celular activada por fluorescencia (FACS). Puede usarse un inmunoensayo basado en monoclonal de dos sitios con anticuerpos monoclonales reactivos para dos epítopos que no interfieren en el polipéptido, pero también puede usarse un ensayo de unión competitiva. Éstos y otros ensayos se describen, entre otros lugares, en Hampton, R. et

al. (1990; Serological Methods, a laboratory Manual, APS Press, St Paul, MN) y Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

Una amplia variedad de marcadores y técnicas de conjugación es conocida para los expertos en la técnica y pueden usarse en varios ensayos de ácidos nucleicos y aminoácidos. Los medios para producir sondas marcadas de hibridación o PCR para detectar secuencias relacionadas con polinucleótidos incluyen oligomarcage, desplazamiento de mella, marcate terminal o amplificación por PCR usando un nucleótido marcado. Alternativamente, las secuencias que codifican los polipéptidos, o cualesquiera fragmentos o variantes de éstos, pueden clonarse en un vector para la producción de una sonda de ARNm. Dichos vectores son conocidos en la técnica, están disponibles comercialmente, y pueden usarse para sintetizar sondas de ARN *in vitro* por adición de una ARN polimerasa apropiada tal como T7, T3, o SP6 y nucleótidos marcados. Estos procedimientos pueden llevarse a cabo usando una variedad de kits disponibles comercialmente Amersham Pharmacia Biotech, Promega, y US Biochemical. Las moléculas informadoras o marcadores adecuados, que pueden usarse para facilitar la detección, incluyen radionúclidos, enzimas, agentes fluorescentes, quimioluminiscentes, o cromogénicos, así como sustratos, cofactores, inhibidores, partículas magnéticas, y semejantes.

Los vectores de expresión o células huésped transformadas con vectores de expresión pueden replicarse en condiciones adecuadas para la expresión y recuperación del polipéptido del cultivo. El cultivo puede comprender componentes para la expresión *in vitro* o *in vivo*. Los componentes de la expresión *in vitro* incluyen aquellos para lisados de reticulocitos de conejo, lisados de *E. coli*, y extractos de germen de trigo, por ejemplo, sistemas Expressway™ o RiPs de Invitrogen, sistemas Genelator™ de iNTRON Biotechnology, sistemas EcoPro™ o STP3™ de Novagen, sistemas acoplados TNT® Quick de Promega, y sistemas EasyXpress de QIAGEN. El polipéptido producido del cultivo puede secretarse o estar contenido intracelularmente, dependiendo de la secuencia y/o el vector usado. En aspectos particulares, los vectores de expresión que codifican un polipéptido de fago pueden diseñarse para contener secuencias señal que dirigen la secreción del polipéptido a través de la membrana de células procariotas o eucariotas.

Otras construcciones pueden incluir un dominio de aminoácido que facilitará la purificación del polipéptido. Dichos dominios incluyen, pero no están limitados a, péptidos quelantes de metales tales como módulos histidina-triptófano (por ejemplo, 6X-HIS (SEQ ID NO: 150)) que permiten la purificación en metales inmovilizados, dominios de proteína A que permiten la purificación en inmunoglobulina inmovilizada, y el dominio utilizado en el sistema de purificación por extensión/afinidad FLAG® (Immunex Corp., Seattle, WA). Las etiquetas de epítipo útiles incluyen 3XFLAG®, HA, VSV-G, V5, HSV, GST, GFP, MBP, GAL4, y β-galactosidasa. Los plásmidos útiles incluyen aquellos que comprenden una etiqueta de biotina (por ejemplo, plásmidos PinPoint™ de Promega), proteína de unión a calmodulina (por ejemplo, plásmidos pCAL de Stratagene), péptido de unión a estreptavidina (por ejemplo, plásmidos InterPlay™ de Stratagene), na etiqueta c-myc o FLAG® (por ejemplo, plásmidos de inmunoprecipitación de Sigma-Aldrich), o una etiqueta de histidina (por ejemplo, plásmidos QIAExpress de QIAGEN).

Para facilitar la purificación, los vectores de expresión pueden incluir secuencias conectoras escindibles tales como aquellas específicas para Factor Xa o enteroquinasa (Invitrogen, San Diego, CA). Por ejemplo, el vector puede incluir uno o más conectores entre el dominio de purificación y el polipéptido. Uno de dichos vectores de expresión proporciona la expresión de una proteína de fusión que comprende un polipéptido de la invención y un ácido nucleico que codifica 6 residuos de histidina precediendo un sitio de escisión de tioredoxina o enteroquinasa. Los residuos de histidina facilitan la purificación en IMAC (cromatografía de afinidad con iones metálicos inmovilizados como se describen en Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281) mientras el sitio de escisión de enteroquinasa proporciona un medio para purificar el polipéptido de la proteína de fusión. Una discusión de vectores que contienen proteínas de fusión se proporciona en Kroll, D. J. et al. (1993; DNA Cell Biol. 12:441-453).

Los anticuerpos descritos aquí pueden producirse usando métodos que son generalmente conocidos en la técnica, por ejemplo, para uso en técnicas de purificación o diagnóstico. En particular, los polipéptidos o polinucleótidos pueden usarse para producir anticuerpos según protocolos generalmente conocidos. Dichos anticuerpos pueden incluir, pero no están limitados a, anticuerpos policlonales, monoclonales, quiméricos, y de cadena única, fragmentos Fab, y fragmentos producidos por una biblioteca de expresión de Fab. Los anticuerpos neutralizantes, (es decir, aquellos que inhiben función) son especialmente preferidos para uso con la invención.

Para la producción de anticuerpos, varios huéspedes incluyendo cabras, conejos, ratas, ratones, seres humanos, y otros, pueden inmunizarse por inyección con un polipéptido, polinucleótido, o cualquier fragmento de éste que tiene propiedades inmunogénicas. Dependiendo de la especie de huésped, pueden usarse varios adyuvantes para incrementar la respuesta inmunológica. Dichos adyuvantes incluyen, pero no están limitados a, geles minerales de Freund tales como hidróxido de aluminio, y sustancias tensioactivas tales como lisolectina, polioles plurónicos, polianiones, péptidos, emulsiones de aceite, hemocianina de lapa, y dinitrofenol. Entre los adyuvantes usados en seres humanos, BCG (bacilo de Calmette-Guerin) y *Corynebacterium parvum* son especialmente preferibles.

Se prefiere que los polipéptidos o fragmentos usados para inducir anticuerpos tengan una secuencia de aminoácidos que comprende al menos cinco aminoácidos y más preferiblemente al menos 10 aminoácidos. También es preferible que sean idénticos a una parte de la secuencia de aminoácidos de la proteína natural, y pueden contener la secuencia de aminoácidos completa de una molécula natural pequeña. Pueden fusionarse cadenas cortas de

aminoácidos con aquellas de otra proteína tal como hemocianina de lapa y producirse el anticuerpo frente a la molécula química.

Los anticuerpos monoclonales pueden prepararse usando cualquier técnica que proporciona la producción de moléculas de anticuerpo por líneas celulares continuas en cultivo. Estas incluyen, pero no están limitadas a, la técnica del hibridoma, la técnica del hibridoma de células B humanas, y la técnica del hibridoma de EBV (Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R. J. et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; Cole, S. P. et al. (1984) *Mol. Cell Biol.* 62:109-120). Los anticuerpos también pueden producirse mediante la inducción de la producción *in vivo* en la población de linfocitos o mediante cribado de bibliotecas de inmunoglobulinas o paneles de reactivos de unión altamente específicos como se describe en la bibliografía (Oriandi, R. et al. (1989) *Proc. Natl. Acad. Sci.* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299).

Además, pueden usarse técnicas para la producción de "anticuerpos quiméricos", por ejemplo, la combinación de genes de anticuerpo para obtener una molécula con la especificidad de antígeno y la actividad biológica apropiadas (Morrison, S. L. et al. (1984) *Proc. Natl. Acad. Sci.* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; Takeda, S. et al. (1985) *Nature* 314:452-454). Alternativamente, pueden adaptarse las técnicas descritas para la producción de anticuerpos de cadena única, usando métodos conocidos en la técnica, para producir anticuerpos de cadena única específicos. Los anticuerpos con especificidad relacionada, pero de distinta composición idiotípica, pueden generarse por intercambio de cadenas de bibliotecas de inmunoglobulinas combinatorias aleatorias (Burton D. R. (1991) *Proc. Natl. Acad. Sci.* 88:11120-3).

Los expertos en la técnica a la que se refiere la invención apreciarán los términos "fragmentos divalentes" y "fragmentos trivalentes". Estos son moléculas que comprenden un dominio variable de cadena pesada (VH) conectado con un dominio variable de cadena ligera (VL) por un conector peptídico corto que es demasiado corto para permitir el emparejamiento entre los dos dominios en la misma cadena. Esto estimula el emparejamiento con los dominios complementarios de una o más otras cadenas y alienta la formación de moléculas dimericas o triméricas con dos o más sitios de unión a antígeno funcionales. Las moléculas de anticuerpo resultantes pueden ser monoespecíficas o multiespecíficas (por ejemplo, biespecíficas en el caso de los fragmentos divalentes). Dichas moléculas de anticuerpo pueden crearse a partir de dos o más anticuerpos usando metodología estándar en la técnica a la que se refiere la invención; por ejemplo, como se describe por Todorovska et al. (Design and application of diabodies, triabodies and tetrabodies for cancer targeting. *J. Immunol. Methods*. 2001 feb 1;248(1-2):47-66).

También pueden generarse fragmentos de anticuerpo que contienen sitios de unión específicos. Por ejemplo, dichos fragmentos incluyen, pero no están limitados a, los fragmentos F(ab')₂ que pueden producirse por digestión con pepsina de la molécula de anticuerpo y los fragmentos Fab que pueden generarse reduciendo los puentes disulfuro de los fragmentos F(ab')₂. Alternativamente, pueden construirse bibliotecas de expresión de Fab para permitir la identificación fácil y rápida de fragmentos Fab monoclonales con la especificidad deseada (Huse, W. D. et al. (1989) *Science* 254:1275-1281).

Pueden usarse varios inmunoensayos para cribado para identificar anticuerpos que tienen especificidad de unión. Numerosos protocolos para ensayos de unión competitiva o inmunoradiométricos usando bien anticuerpos policlonales o monoclonales con especificidades establecidas son muy conocidos en la técnica. Dichos inmunoensayos típicamente implican la medida de la formación de un complejo entre un polipéptido o polinucleótido y su anticuerpo específico. Se prefiere un inmunoensayo basado en monoclonal de dos sitios que utiliza anticuerpos monoclonales reactivos frente a dos epítopos que no interfieren, pero también puede emplearse un ensayo de unión competitiva (Maddox, supra).

Los polipéptidos de fago descritos en la presente memoria tienen la capacidad de tomar como diana, permeabilizar, y/o inhibir células y también son útiles como moléculas transportadoras para la administración de moléculas inhibitorias adicionales en células microbianas. La química para el acoplamiento de compuestos a aminoácidos está muy desarrollada y varios tipos de moléculas diferentes podrían ligarse a los polipéptidos. Los métodos de acoplamiento más comunes se basan en la presencia de grupos libres amino (alfa-amino o Lys), sulfhidrilo (Cys), o ácido carboxílico (Asp, Glu, o alfa-carboxilo). Los métodos de acoplamiento pueden usarse para ligar el polipéptido al inhibidor celular mediante el residuo carboxi- o amino-terminal. En algunos casos, una secuencia incluye múltiples residuos que pueden reaccionar con la química elegida. Esto puede usarse para producir multímeros, que comprenden más de un inhibidor celular. Alternativamente, el polipéptido puede acortarse o elegirse de manera que los residuos reactivos estén localizados bien en el extremo amino o carboxilo de la secuencia.

Por ejemplo, una molécula informadora tal como fluoresceína puede incorporarse específicamente en un residuo de lisina (Ono et al., 1997) usando *N*-α-Fmoc- *N*ε-1-(4,4-dimetil-2,6 dioxociclohex-1-iliden-3-metilbutil)-L-lisina durante la síntesis del polipéptido. Después de la síntesis, los 5- y 6-carboxifluoresceína succinimidil ésteres pueden acoplarse después de eliminar 4,4-dimetil-2,6 dioxociclohex-1-iliden por tratamiento con hidrazina. Por lo tanto, el acoplamiento de una molécula inhibitoria al polipéptido de fago puede conseguirse por inclusión de un residuo de lisina en la secuencia del polipéptido, después reacción con un inhibidor celular derivatizado adecuadamente.

También puede usarse EDC (hidrocloruro de 1-etil-3-(3-dimetilaminopropil) carbodiimida) o el método de acoplamiento de carbodiimida. Las carbodiimidas pueden activar los grupos carboxílicos de la cadena lateral del ácido aspártico y glutámico, así como el grupo carboxilo terminal para hacerlos sitios reactivos para el acoplamiento con aminas primarias. Los polipéptidos activados se mezclan con el inhibidor celular para producir el conjugado final.

5 Si el inhibidor celular se activa en primer lugar, el método de EDC acoplará el inhibidor celular a través de la alfa amina N-terminal y posiblemente a través de la amina en la cadena lateral de Lys, si está presente en la secuencia.

El éster de m-maleimidobenzoil-N-hidroxisuccinimida (MBS) es un reactivo heterobifuncional que puede usarse para unir polipéptidos a inhibidores celulares a través de cisteínas. El acoplamiento tiene lugar con el grupo tiol de residuos de cisteína. Si la secuencia elegida no contiene Cys es común poner un residuo de Cys en el extremo N o C para obtener la unión altamente controlada del polipéptido al inhibidor celular. Para propósitos de síntesis, puede

10 ayudar poner la cisteína en el extremo N del polipéptido. MBS es particularmente idóneo para uso con la presente invención.

El glutaraldehído puede usarse como un reactivo de acoplamiento bifuncional que une dos compuestos a través de sus grupos amino. El glutaraldehído proporciona un espaciador altamente flexible entre el polipéptido y el inhibidor celular para una presentación favorable. El glutaraldehído es un compuesto muy reactivo y reaccionará con Cys, Tyr, e His en un grado limitado. El método de acoplamiento con glutaraldehído es particularmente útil cuando un polipéptido sólo contiene un único grupo amino libre en su extremo amino. Si el polipéptido contiene más de un grupo amino libre, pueden formarse grandes complejos multiméricos.

15 El glutaraldehído puede usarse como un reactivo de acoplamiento bifuncional que une dos compuestos a través de sus grupos amino. El glutaraldehído proporciona un espaciador altamente flexible entre el polipéptido y el inhibidor celular para una presentación favorable. El glutaraldehído es un compuesto muy reactivo y reaccionará con Cys, Tyr, e His en un grado limitado. El método de acoplamiento con glutaraldehído es particularmente útil cuando un polipéptido sólo contiene un único grupo amino libre en su extremo amino. Si el polipéptido contiene más de un grupo amino libre, pueden formarse grandes complejos multiméricos.

En un aspecto, los polipéptidos de la invención pueden fusionarse (por ejemplo, por clonación en marco) o unirse (por ejemplo, por acoplamiento químico) a inhibidores celulares tales como agentes antimicrobianos. Incluidos entre éstos están péptidos antimicrobianos, por ejemplo, proteína bactericida/que incrementa la permeabilidad, proteínas antimicrobianas catiónicas, lisozimas, lactoferrinas, y catelicidinas (por ejemplo, de neutrófilos; véase, por ejemplo, Hancock y Chappie, 1999, Antimicrob. Agents Chemother.43:1317-1323; Ganz y Lehrer, 1997, Curr. Opin. Hematol. 4:53- 58; Hancock et al., 1995, Adv. Microb. Physiol. 37:135-175). Los péptidos antimicrobianos incluyen además defensinas (por ejemplo, de células epiteliales o neutrófilos) y proteínas microbicidas de plaquetas (véase, por ejemplo, Hancock y Chappie, 1999, Antimicrob. Agents Chemother.43:1317-1323). Los péptidos antimicrobianos adicionales incluyen, pero no están limitados a, gramicidina S, bacitracina, polimixina B, taquiplesina, batenecina (por ejemplo, batenecina de ganado), ranalexina, cecropina A, indolicidina (por ejemplo, indolicidina de ganado), y nisina (por ejemplo, nisina bacteriana).

20 (por ejemplo, por acoplamiento químico) a inhibidores celulares tales como agentes antimicrobianos. Incluidos entre éstos están péptidos antimicrobianos, por ejemplo, proteína bactericida/que incrementa la permeabilidad, proteínas antimicrobianas catiónicas, lisozimas, lactoferrinas, y catelicidinas (por ejemplo, de neutrófilos; véase, por ejemplo, Hancock y Chappie, 1999, Antimicrob. Agents Chemother.43:1317-1323; Ganz y Lehrer, 1997, Curr. Opin. Hematol. 4:53- 58; Hancock et al., 1995, Adv. Microb. Physiol. 37:135-175). Los péptidos antimicrobianos incluyen además defensinas (por ejemplo, de células epiteliales o neutrófilos) y proteínas microbicidas de plaquetas (véase, por ejemplo, Hancock y Chappie, 1999, Antimicrob. Agents Chemother.43:1317-1323). Los péptidos antimicrobianos adicionales incluyen, pero no están limitados a, gramicidina S, bacitracina, polimixina B, taquiplesina, batenecina (por ejemplo, batenecina de ganado), ranalexina, cecropina A, indolicidina (por ejemplo, indolicidina de ganado), y nisina (por ejemplo, nisina bacteriana).

También incluidos como agentes antimicrobianos están ionóforos, que facilitan la transmisión de un ion, (tal como sodio), a través de una barrera lipídica tal como una membrana celular. Dos compuestos ionóforos particularmente idóneos para esta invención son el RUMENSIN™ (Eli Lilly) y Lasalocid (Hoffman LaRoche). Otros ionóforos incluyen, pero no están limitados a, salinomycin, avoparcina, aridicina, y actaplanina. Otros agentes antimicrobianos incluyen penicilina, Monensina™ y azitromicina, metronidazol, estreptomycin, kanamicina, y penicilina, así como, generalmente, β-lactamas, aminoglicósidos, macrólidos, cloranfenicol, novobiocina, rifampina, y fluoroquinolonas (véase, por ejemplo, Horn et al., 2003, Applied Environ. Microbiol. 69:74-83; Eckburg et al., 2003, Infection Immunity 71 :591-596; Gijzen et al., 1991, Applied Environ. Microbiol. 57:1630-1634; Bonelo et al., 1984, FEMS Microbiol. Lett. 21:341-345; Huser et al., 1982, Arch. Microbiol. 132:1-9; Hilpert et al., 1981, Zentbl. Bakteriell. Mikrobiol. Hyg. 1 Abt Orig. C 2:21-31).

30 También incluidos como agentes antimicrobianos están ionóforos, que facilitan la transmisión de un ion, (tal como sodio), a través de una barrera lipídica tal como una membrana celular. Dos compuestos ionóforos particularmente idóneos para esta invención son el RUMENSIN™ (Eli Lilly) y Lasalocid (Hoffman LaRoche). Otros ionóforos incluyen, pero no están limitados a, salinomycin, avoparcina, aridicina, y actaplanina. Otros agentes antimicrobianos incluyen penicilina, Monensina™ y azitromicina, metronidazol, estreptomycin, kanamicina, y penicilina, así como, generalmente, β-lactamas, aminoglicósidos, macrólidos, cloranfenicol, novobiocina, rifampina, y fluoroquinolonas (véase, por ejemplo, Horn et al., 2003, Applied Environ. Microbiol. 69:74-83; Eckburg et al., 2003, Infection Immunity 71 :591-596; Gijzen et al., 1991, Applied Environ. Microbiol. 57:1630-1634; Bonelo et al., 1984, FEMS Microbiol. Lett. 21:341-345; Huser et al., 1982, Arch. Microbiol. 132:1-9; Hilpert et al., 1981, Zentbl. Bakteriell. Mikrobiol. Hyg. 1 Abt Orig. C 2:21-31).

Los inhibidores particularmente útiles son compuestos que bloquean o interfieren con la metanogénesis, incluyendo ácido bromoetanosulfónico, por ejemplo, ácido 2-bromoetanosulfónico (BES) o una sal de éste, por ejemplo, una sal de sodio. El molibdato de sodio (Mo) es un inhibidor de la reducción de sulfato, y puede usarse con ácido bromoetanosulfónico. Otros compuestos anti-metanogénesis incluyen, pero no están limitados a, nitrato, formato, fluoruro de metilo, cloroformo, hidrato de cloral, sulfito de sodio, etileno e hidrocarburos insaturados, acetileno, ácidos grasos tal como ácido linoleico y cis-oleico, ácidos grasos saturados tal como ácido behénico y esteárico, y, también lumazina (por ejemplo, 2,4-pteridinadiona). Los compuestos adicionales incluyen 3-bromopropanosulfonato (BPS), ácido propinoico, y 2-butinoato de etilo.

40 Los inhibidores particularmente útiles son compuestos que bloquean o interfieren con la metanogénesis, incluyendo ácido bromoetanosulfónico, por ejemplo, ácido 2-bromoetanosulfónico (BES) o una sal de éste, por ejemplo, una sal de sodio. El molibdato de sodio (Mo) es un inhibidor de la reducción de sulfato, y puede usarse con ácido bromoetanosulfónico. Otros compuestos anti-metanogénesis incluyen, pero no están limitados a, nitrato, formato, fluoruro de metilo, cloroformo, hidrato de cloral, sulfito de sodio, etileno e hidrocarburos insaturados, acetileno, ácidos grasos tal como ácido linoleico y cis-oleico, ácidos grasos saturados tal como ácido behénico y esteárico, y, también lumazina (por ejemplo, 2,4-pteridinadiona). Los compuestos adicionales incluyen 3-bromopropanosulfonato (BPS), ácido propinoico, y 2-butinoato de etilo.

Se incluyen además como agentes antimicrobianos enzimas líticas, incluyendo lisozima de fago, endolisina, lisozima, lisina, lisina de fago, muralisina, muramidasa, y virolisina. Las enzimas útiles presentan la capacidad de hidrolizar enlaces específicos en la pared celular bacteriana. Las enzimas líticas particulares incluyen, pero no están limitadas a, glucosaminidasas, que hidrolizan los enlaces glicosídicos entre los azúcares amino (por ejemplo, ácido N-acetilmurámico y N-acetilglucosamina) del peptidoglicano, amidasas, que escinden la unión N-acetilmuramoyl-L-alanina amida entre la cadena de glicano y el péptido de entrecruzamiento, y endopeptidasas, que hidrolizan la unión interpeptídica (por ejemplo, cisteína endopeptidasas) y endoisopeptidasas que atacan pseudomureína de metanógenos de la familia *Methanobacteriaceae*.

50 Se incluyen además como agentes antimicrobianos enzimas líticas, incluyendo lisozima de fago, endolisina, lisozima, lisina, lisina de fago, muralisina, muramidasa, y virolisina. Las enzimas útiles presentan la capacidad de hidrolizar enlaces específicos en la pared celular bacteriana. Las enzimas líticas particulares incluyen, pero no están limitadas a, glucosaminidasas, que hidrolizan los enlaces glicosídicos entre los azúcares amino (por ejemplo, ácido N-acetilmurámico y N-acetilglucosamina) del peptidoglicano, amidasas, que escinden la unión N-acetilmuramoyl-L-alanina amida entre la cadena de glicano y el péptido de entrecruzamiento, y endopeptidasas, que hidrolizan la unión interpeptídica (por ejemplo, cisteína endopeptidasas) y endoisopeptidasas que atacan pseudomureína de metanógenos de la familia *Methanobacteriaceae*.

Los polipéptidos codificados por ORF 2058 o ORF 2055, descritos con detalle en la presente memoria y más adelante, son útiles como enzimas líticas específicas de metanógenos del rumen. Las enzimas nativas pueden prepararse a partir de células de *M. ruminantium* lisadas con φmru frescas. Alternativamente, ORF 2058 o ORF 2055 pueden clonarse en un vector de expresión y expresarse en un huésped heterólogo tal como *Escherichia coli*. Esto se ha conseguido previamente con PeiP y PeiW y se mostró que las proteínas recombinantes eran activas frente a

60 Los polipéptidos codificados por ORF 2058 o ORF 2055, descritos con detalle en la presente memoria y más adelante, son útiles como enzimas líticas específicas de metanógenos del rumen. Las enzimas nativas pueden prepararse a partir de células de *M. ruminantium* lisadas con φmru frescas. Alternativamente, ORF 2058 o ORF 2055 pueden clonarse en un vector de expresión y expresarse en un huésped heterólogo tal como *Escherichia coli*. Esto se ha conseguido previamente con PeiP y PeiW y se mostró que las proteínas recombinantes eran activas frente a

las paredes celulares de *Methanothermobacter* en condiciones reductoras (Luo et al., 2002). Las enzimas líticas ORF 2058 o ORF 2055 o cualquier otra enzima lítica pueden usarse en composiciones, por ejemplo, como un aditivo alimentario para rumiantes o pueden incorporarse en una cápsula o dispositivo de bolo de liberación lenta para la administración durante un periodo de tiempo más largo en el rumen. Las enzimas líticas pueden usarse bien en combinación o secuencialmente con otro u otros inhibidores de metanógenos para evitar la adaptación de los metanógenos del huésped y la resistencia a las enzimas. También pueden usarse mutaciones aleatorias y/o dirigidas en las enzimas para evitar la adaptación. Los componentes del cambio lítico/lisogénico (por ejemplo, ORF 1981 y ORF 1983-ORF 1986) pueden usarse de una manera similar a la de las enzimas líticas.

Adicionalmente, se incluyen PNA como agentes antimicrobianos. Los PNA son híbridos péptido-ácido nucleico, en los que el núcleo fosfato se ha reemplazado por un núcleo aciclico y neutro compuesto a partir de unidades de *N*-(2-aminoetil)-glicina (véase, por ejemplo, Eureka Bioscience Collection. PNA and Oligonucleotide Inhibitors of Human Telomerase. G. Gavory y S. Balasubramanian, Landes Bioscience, 2003). Las bases A, G, T, C están unidas al nitrógeno amino en el núcleo a través de uniones metilencarbonilo (P.E. Nielsen et al., Science 1991. 254: 1497-1500; M. Egholm et al., Nature 1993. 365: 566-568). Los PNA se unen a secuencias complementarias con alta especificidad, y mayor afinidad respecto a ADN o ARN análogos (M. Egholm et al., supra). Los híbridos PNA/ADN o PNA/ARN también presentan una mayor estabilidad térmica comparados con los dúplex correspondientes ADN/ADN o ADN/ARN (M. Egholm et al., supra). Los PNA también poseen una alta estabilidad química y biológica, debido al núcleo amida no natural que no es reconocido por nucleasas o proteasas (V. Demidov et al., Biochem Pharmacol 1994. 48: 1310-1313). Típicamente, los PNA tienen una longitud de al menos 5 bases, e incluyen una lisina terminal. Los PNA pueden pegarse para prolongar adicionalmente su vida (Nielsen, P. E. et al. (1993) Anticancer Drug Des. 8:53-63).

En un aspecto particular, los polipéptidos de la invención pueden fusionarse (por ejemplo, por clonación en marco) o unirse (por ejemplo, por acoplamiento químico) con inhibidores celulares tales como anticuerpos o fragmentos de éstos. Los anticuerpos o fragmentos de anticuerpo pueden estar dirigidos a células microbianas, o particularmente células metanógenas, o uno o más componentes celulares. Por ejemplo, las proteínas de la superficie celular, por ejemplo, receptores extracelulares, pueden ser una diana. Se incluyen moléculas de inmunoglobulina y partes inmunológicamente activas de moléculas de inmunoglobulina (Ig), es decir, moléculas que contiene un sitio de unión a antígeno que se une específicamente (inmunorreacciona con) un antígeno.

Los polipéptidos de la invención encuentran un uso particular en tomar como diana una célula microbiana, en particular, una célula metanógena. En determinados aspectos, los polipéptidos pueden usarse para asociarse con o unirse a la pared o membrana celular, permeabilizar la célula, y/o inhibir el crecimiento o replicación de la célula. Como tales, los polipéptidos pueden usarse para la unión temporal o prolongada a la célula, o para penetrar la pared o membrana celular y/o acumularse en el entorno intracelular. Se entiende que los polipéptidos de fago, así como los polinucleótidos, vectores de expresión, células huésped, y anticuerpos de la descripción correspondientes, pueden usarse para tomar como diana varios microbios, por ejemplo, *Methanobrevibacter ruminantium*, que es el metanógeno principal en los rumiantes, y *Methanobrevibacter smithii*, que es el metanógeno principal en los seres humanos. Para efectuar el direccionamiento, la célula microbiana puede ponerse en contacto con el polipéptido de fago como aislado de una o más fuentes naturales, o producido por vectores de expresión y/o células huésped, o química sintética o semi-sintética como se describe con detalle en la presente memoria. Para una permeabilización potenciada, el polipéptido puede fusionarse o unirse a una o más secuencias señal (secuencia consenso predicha: [ML]K[KKK][K]{0,1}X{0,9}[IL][IFL][IL][IL][IS][LIA]X{0,4}[LIVF][LIAV][LI][LV][LIAV][ILFV][LIVF][SAL][ILV][GSA][AS][VAI][S A][A (SEQ ID NO: 151), véase la FIG. 6). Véase también Perez-Bercoff, A., Koch, J. y Bürglin, T.R. (2006) LogoBar: bar graph visualization of protein logos with gaps. Bioinformatics 22, 112-114. En aspectos particulares, el polipéptido se administra a sujetos como una composición descrita con detalle en la presente memoria, por ejemplo, mediante el uso de un dispositivo de liberación lenta para rumiantes.

En determinadas realizaciones, el polipéptido se fusiona o une a un inhibidor celular, por ejemplo, un compuesto anti-metanogénesis (por ejemplo, ácido bromoetanosulfónico), un anticuerpo o fragmento de anticuerpo, enzima lítica, ácido nucleico peptídico, péptido antimicrobiano, u otro antibiótico. El polipéptido-inhibidor se administra a sujetos como una composición para inhibir el crecimiento y/o replicación de células microbianas, en particular, células metanógenas. La composición comprende, por ejemplo: a) un fago aislado, partícula de fago, genoma de fago, o alteración, fragmento, variante, o derivado de éstos; b) un polipéptido de fago aislado, o una alteración, fragmento, variante, o derivado de éste; c) un polinucleótido aislado, o una alteración, fragmento, variante, o derivado de éste; d) un vector de expresión que comprende este polinucleótido; o e) una célula huésped que comprende este vector de expresión. Las composiciones descritas aquí pueden envasarse específicamente como parte de kits para tomar como diana, permeabilizar, y/o inhibir células microbianas, especialmente células metanógenas, según los métodos descritos. Los kits comprenden al menos una composición como se muestra en la presente memoria e instrucciones para uso en la toma como diana o permeabilización de células, o inhibición del crecimiento o replicación celular, para metanógenos u otros microbios.

Como una realización adicional, la descripción se refiere a una composición farmacéutica en conjunción con un vehículo farmacéuticamente aceptable, para uso con cualquiera de los métodos discutidos anteriormente. Dichas composiciones farmacéuticas pueden comprender un polipéptido de fago, en combinación con un inhibidor celular. Alternativamente, las composiciones farmacéuticas pueden comprender un vector de expresión o célula huésped

como se describe con detalle en la presente memoria. Las composiciones pueden administrarse solas o en combinación con al menos un otro agente, tal como un compuesto estabilizador, que puede administrarse en cualquier vehículo farmacéutico estéril, biocompatible, incluyendo, pero no limitado a, disolución salina, disolución salina tamponada, dextrosa, y agua. Las composiciones pueden administrarse a un sujeto solas, o en combinación con otros agentes, fármacos (por ejemplo, fármacos antimicrobianos), u hormonas.

Además de los ingredientes activos, estas composiciones farmacéuticas pueden contener vehículos farmacéuticamente aceptables adecuados que comprenden excipientes y auxiliares que facilitan el procesamiento de los compuestos activos en preparaciones que pueden usarse farmacéuticamente. Los detalles adicionales sobre las técnicas de formulación y administración pueden encontrarse en la última edición de Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA). Las composiciones farmacéuticas utilizadas en esta invención pueden administrarse por cualquier número de rutas incluyendo, pero no limitado a, medios orales, intravenosos, intramusculares, intra-arteriales, intramedulares, intratecales, intraventriculares, transdérmicos, subcutáneos, intraperitoneales, intranasales, enterales, tópicos, sublinguales, o rectales.

Las composiciones farmacéuticas para administración oral pueden formularse usando vehículos farmacéuticamente aceptables muy conocidos en la técnica en dosificaciones adecuadas para administración oral. Dichos vehículos permiten que las composiciones farmacéuticas se formulen como comprimidos, píldoras, grageas, cápsulas, líquidos, geles, jarabes, suspensiones de sólidos, suspensiones, y semejantes, para ingestión por el sujeto. Las preparaciones farmacéuticas para uso oral pueden obtenerse mediante la combinación de los compuestos activos con excipiente sólido, opcionalmente triturando una mezcla resultante, y procesando la mezcla de gránulos, después de añadir auxiliares adecuados, si se desea, para obtener comprimidos o núcleos de grageas. Los excipientes adecuados son rellenos de carbohidratos o proteínas, tales como azúcares, incluyendo lactosa, sacarosa, manitol, o sorbitol; almidón de maíz, trigo, arroz, patata, u otras plantas; celulosa, tal como metil celulosa, hidroxipropilmetilcelulosa, o carboximetilcelulosa de sodio; gomas incluyendo arábica y de tragacanto; y proteínas tal como gelatina y colágeno. Si se desea, pueden añadirse agentes disgregantes o solubilizantes, tal como polivinil pirrolidona entrecruzada, agar, ácido algínico, o una sal de éste, tal como alginato de sodio.

Las preparaciones farmacéuticas que pueden usarse oralmente incluyen cápsulas de ajuste a presión hechas de gelatina, así como cápsulas selladas blandas hechas de gelatina y un recubrimiento, tal como glicerol o sorbitol. Las cápsulas de ajuste a presión pueden contener ingredientes activos mezclados con un relleno o aglutinantes, tales como lactosa o almidones, lubricantes, tales como talco o estearato de magnesio, y, opcionalmente, estabilizadores. En las capsulas blandas, los compuestos activos pueden estar disueltos o suspendidos en líquidos adecuados, tales como aceites grasos, líquido, o polietilen glicol líquido con o sin estabilizadores. Los núcleos de grageas pueden usarse en conjunción con recubrimientos adecuados, tales como disoluciones concentradas de azúcar, que también pueden contener goma arábica, talco, polivinilpirrolidona, gel carbopol, polietilen glicol, y/o dióxido de titanio, disoluciones de laca, y disolventes orgánicos adecuados o mezclas de disolventes. Pueden añadirse materiales colorantes o pigmentos a los comprimidos o recubrimientos de grageas para la identificación del producto o para caracterizar la cantidad de compuesto activo, es decir, dosificación.

Las formulaciones farmacéuticas adecuadas para administración parenteral pueden formularse en disoluciones acuosas, preferiblemente en tampones fisiológicamente compatibles tales como disolución de Hanks, disolución de Ringer, o disolución salina fisiológicamente tamponada. Las suspensiones acuosas para inyección pueden contener sustancias que incrementan la viscosidad de la suspensión, tales como carboximetil celulosa de sodio, sorbitol, o dextrano. Adicionalmente, las suspensiones de los compuestos activos pueden prepararse como suspensiones aceitosas apropiadas para inyección. Los disolventes o vehículos lipofílicos adecuados incluyen aceites grasos tales como aceite de sésamo, o ésteres de ácidos grasos sintéticos, tales como oleato de etilo o triglicéridos o liposomas. También pueden usarse polímeros de amino policatiónicos no lipídicos para la administración. Opcionalmente, la suspensión también puede contener estabilizadores adecuados o agentes que incrementan la solubilidad de los compuestos para permitir la preparación de disoluciones altamente concentradas. Para la administración tópica o nasal, se usan en la formulación penetrantes apropiados para la barrera particular que se va a permear. Dichos penetrantes son conocidos generalmente en la técnica.

Las composiciones farmacéuticas de la presente descripción pueden fabricarse de una manera que es conocida en la técnica, por ejemplo, mediante procesos convencionales de mezclado, disolución, granulación, preparación de grageas, levigación, emulsión, encapsulación, atrapamiento, o liofilización. La composición farmacéutica puede proporcionarse como una sal y puede formarse con muchos ácidos, incluyendo, pero no limitado a, clorhídrico, sulfúrico, acético, láctico, tartárico, málico, succínico, etc. Las sales tienden a ser más solubles en disolventes acuosos u otros protónicos de lo que lo son las formas de base libre correspondientes. En otros casos, la preparación preferida puede ser un polvo liofilizado que puede contener cualquiera o todos de los siguientes: 1-50 mM histidina, 0,1%-2% sacarosa, y 2-7% manitol, a un intervalo de pH de 4,5 a 5,5, que se combina con un tampón antes del uso. Después de que se hayan preparado las composiciones farmacéuticas, pueden ponerse en un contenedor apropiado y etiquetarse para el tratamiento de una afección indicada. Para la administración de una composición de la descripción, dicho etiquetado incluirá la cantidad, frecuencia, y método de administración.

Las composiciones farmacéuticas adecuadas para uso en la presente descripción incluyen composiciones en las que los ingredientes activos están contenidos en una cantidad efectiva para conseguir el propósito pretendido. Para

cualquier compuesto, la dosis terapéuticamente efectiva puede estimarse inicialmente bien en ensayos celulares, por ejemplo, en células microbianas, o en particular, en células metanógenas, o en modelos animales, habitualmente ratones, conejos, perros, o cerdos, o en especies de rumiantes, tales como ovejas, ganado, ciervos, y cabras. El modelo animal también puede usarse para determinar el intervalo apropiado de concentración y la ruta de administración. Dicha información puede usarse entonces para determinar las dosis útiles y las rutas para administración. Las cantidades de dosificación normales pueden variar de 0,1 a 100.000 microgramos, hasta una dosis total de aproximadamente 1 g, o más, dependiendo de la ruta de administración. La guía respecto a las dosificaciones particulares y los métodos de administración se proporciona en la bibliografía y está disponible generalmente para los expertos en la técnica. Los expertos en la técnica emplearán formulaciones diferentes para polinucleótidos que para polipéptidos. De forma similar, la administración de polinucleótidos o polipéptidos será específica para células, condiciones, localizaciones particulares, etc.

Los terapéuticos basados en fago se conocen, y los métodos para la fabricación de dichas composiciones están publicados en la técnica. Los terapéuticos de fago se han descrito, por ejemplo, para tomar como diana *Staphylococcus* (por ejemplo, *S. aureus*), *Pseudomonas* (por ejemplo, *P. aeruginosa*), *Escherichia* (por ejemplo, *E. coli*), *Klebsiella* (por ejemplo, *K. ozaenae*, *K. rhinoscleromatis* *scleromatis* y *K. pneumonia*), *Proteus*, *Salmonella*, *Shigella* (véase, por ejemplo, Carlton, R.M. (1999). *Archivum Immunologiae et Therapiae Experimentalis*, 47: 267-274; Liu, J. et al. (2004). *Nat. Biotechnol.* 22, 185-191; Projan, S. (2004). *Nat. Biotechnol.* 22, 167-168; Sulakvelidze, A., Alavidze, Z. y Morris, J. G. (2001). *Antimicrobial Agents and Chemotherapy*, 45 (3): 649-659; Weber-Dabrowska, Mulczyk, M. y Gorski, A. (2000). *Archivum Immunologiae et Therapiae Experimentalis*, 48: 547-551. Las terapias de fago tienen ventajas inherentes sobre los anti-microbianos tradicionales, ya que los fagos son altamente específicos y no afectan la microflora normal del cuerpo humano; los fagos no infectan células eucariotas, y no tienen efectos secundarios graves conocidos; los fagos se pueden localizar en el sitio de la infección; y los fagos se pueden replicar exponencialmente, de manera que los tratamientos requieren sólo una pequeña dosis y tienen generalmente un coste bajo (véase, por ejemplo, Sulakvelidze et al., supra). Para una revisión actual, véase Fischetti VA, Nelson D, Schuch R. Reinventing phage therapy: are the parts greater than the sum? *Nat Biotechnol.* 2006 dic;24(12): 1508-11.

Los terapéuticos basados en péptidos y polipéptidos también se han descrito, por ejemplo, para denileuquina, difitox, octreótido, vapreótido, lanreótido, la serie de péptidos RC-3940, decapeptilo, luprón, zoladex, cetorelix (véase, por ejemplo, Lu et al., 2006, *AAPS J* 8:E466-472), hemocidinas, estafopainas (véase, por ejemplo, Dubin et al., 2005, *Acta Biochemica Polonica*, 52:633-638), así como indolicidina, defensinas, lantibióticos, microcidina B17, histatinas, y maganina (véase, por ejemplo, Yeaman y Yount, 2003, *Pharmacol Rev* 55:27-55). La guía general para terapéuticos peptídicos y polipeptídicos también puede encontrarse en Degim et al., 2007, *Curr Pharm Des* 13:99-117 y Shai et al., 2006, *Curr Prot Pept Sci*, 7:479-486. Los fármacos basados en péptidos aprobados recientemente incluyen Hematide™ (agente estimulante de la eritropoyesis basado en péptido sintético, Affymax, Inc.), Exenatide (exendina-4 sintética, Amylin/Eli Lilly), Natreco (nesiritida, péptido natriurético, Scios), Plenaxis (abarelix, Praecis Pharmaceuticals), y SecreFlo (secretina, Repligen).

La dosificación exacta será determinada por el médico, a la luz de factores relacionados con el sujeto que requiere el tratamiento. La dosificación y administración se ajustan para proporcionar niveles suficientes del agente activo o para mantener el efecto deseado. Los factores que pueden tenerse en cuenta incluyen la gravedad del estado de enfermedad, salud general del sujeto, edad, peso, y sexo del sujeto, dieta, tiempo, y frecuencia de la administración, combinaciones de fármacos, sensibilidades de reacción, y tolerancia/respuesta a la terapia. Las composiciones farmacéuticas de actuación prolongada pueden administrarse cada 3 a 4 días, cada semana, o una vez cada dos semanas dependiendo de la vida media y velocidad de aclaramiento de la formulación particular.

Particularmente útiles para las composiciones de la descripción (por ejemplo, composiciones farmacéuticas) son fórmulas o mecanismos de liberación lenta. Por ejemplo, los dispositivos intra-ruminales incluyen, pero no están limitados a, Time Capsule™ Bolus range por Agri-Feeds Ltd., Nueva Zelanda, desarrollado originalmente en AgResearch Ltd., Nueva Zelanda, como se describe en WO 95/19763 y NZ 278977, y CAPTEC por Nufarm Health & Sciences, una división de Nufarm Ltd., Auckland, Nueva Zelanda, como se describen en AU 35908178, PCT/AU81/100082, y Laby et al., 1984, *Can. J. Anim. Sci.* 64 (Supl.), 337-8. Como un ejemplo particular, el dispositivo puede incluir un muelle y émbolo que fuerza a la composición frente a un agujero en el extremo de un cilindro.

Como una realización adicional, la descripción se refiere a una composición para un suplemento de agua, por ejemplo, composición de rociado, o suplemento alimenticio, por ejemplo, componente de alimento de rumiantes, para uso con cualquiera de los métodos discutidos anteriormente. En aspectos particulares, el suplemento alimenticio comprende al menos un material vegetal que es comestible, y un péptido o polipéptido de la invención. Alternativamente, el suplemento alimenticio comprende al menos un material vegetal que es comestible, y un polipéptido o péptido, o un polinucleótido que codifica un péptido o polipéptido descrito en la presente memoria, por ejemplo, como un vector de expresión o célula huésped que comprende el vector de expresión. En particular, la composición incluye además un inhibidor celular, como fusionado o unido a la secuencia resultante. El material vegetal preferido incluye cualquiera de heno, hierba, grano, o harina de maíz, por ejemplo, heno de legumbre, heno de hierba, ensilaje de maíz, ensilaje de hierba, ensilaje de legumbres, grano de maíz, avenas, cebada, grano destilado, grano de cerveza, harina de soja, y harina de semilla de algodón. En particular, el ensilaje de hierba es útil como una composición alimenticia para rumiantes. El material de planta puede modificarse genéticamente para

contener uno o más componentes de la invención, por ejemplo, uno o más polipéptidos o péptidos, polinucleótidos, o vectores.

En otra realización, los anticuerpos que se producen frente a los polipéptidos o polinucleótidos de la descripción pueden usarse para determinar la presencia de microbios, especialmente metanógenos, o en ensayos para monitorizar los niveles de dichos microbios. Los anticuerpos útiles para propósitos de diagnóstico pueden prepararse de la misma manera que la descrita anteriormente. Los ensayos de diagnóstico incluyen métodos que utilizan el anticuerpo y un marcador para detectar un polipéptido en los fluidos del cuerpo humano o extractos de células o tejidos: Los anticuerpos pueden usarse con o sin modificación, y pueden marcarse mediante la unión a ellos, bien covalentemente o no covalentemente, de una molécula informadora. Una amplia variedad de moléculas informadoras que se conocen en la técnica puede usarse, varias de las cuales se han descrito anteriormente.

Una variedad de protocolos para medir los niveles de un polipéptido o polinucleótido son conocidos en la técnica (por ejemplo, ELISA, RIA, FACS, y transferencias, tales como transferencias Southern, Northern, Western), y proporcionan una base para determinar la presencia o niveles de un microbio, especialmente un metanógeno. Los niveles normales o estándar se establecen mediante la combinación de fluidos corporales o extractos de células tomados de sujetos normales, por ejemplo, seres humanos o rumiantes normales, con el anticuerpo en condiciones adecuadas para la formación de complejos. La cantidad de formación de complejos estándar puede cuantificarse por varios métodos, pero preferiblemente por medios fotométricos. Las cantidades de polipéptido o polinucleótido expresado en muestras del sujeto, control, y tratadas (por ejemplo, muestras de sujetos tratados) se comparan con los valores estándar. La desviación entre los valores estándar y del sujeto establece los parámetros para determinar la presencia o niveles del microbio.

En una realización particular de la descripción, los polinucleótidos pueden usarse para propósitos de diagnóstico usando técnicas particulares de hibridación y/o amplificación. Los polinucleótidos que pueden usarse incluyen oligonucleótidos, moléculas complementarias de ARN y ADN, y PNA. Los polinucleótidos pueden usarse para detectar y cuantificar la expresión génica en muestras en las que la expresión puede correlacionarse con la presencia o niveles de un microbio. El ensayo de diagnóstico puede usarse para distinguir entre la ausencia, presencia, y alteración de los niveles de microbio, y para monitorizar los niveles durante la intervención terapéutica.

En un aspecto, la hibridación con sondas de PCR puede usarse para identificar secuencias de ácido nucleico, especialmente secuencias genómicas, que codifican los polipéptidos de la invención. La especificidad de la sonda, ya se haya hecho a partir de una región altamente específica, por ejemplo, 10 nucleótidos únicos en la región reguladora 5', o una región menos específica, por ejemplo, en la región codificadora 3', y la astringencia de la hibridación o amplificación (máxima, alta, intermedia, o baja) determinará si la sonda identifica sólo secuencias naturales, alelos, o secuencias relacionadas. Las sondas también pueden usarse para la detección de secuencias relacionadas, y deben contener preferiblemente al menos 50% de los nucleótidos de cualquiera de las secuencias codificadoras. Las sondas de hibridación de la invención objeto pueden ser ADN o ARN y derivar de la secuencia de nucleótidos de SEQ ID NO: 74-142, o complementos, o secuencias modificadas de éstas, o de secuencias genómicas incluyendo promotor, elementos potenciadores, e intrones de la secuencia natural.

Los medios para producir sondas de hibridación específicas para ADN incluyen la clonación de secuencias de ácido nucleico en vectores para la producción de sondas de ARNm. Dichos vectores son conocidos en la técnica, están disponibles comercialmente, y pueden usarse para sintetizar sondas de ARN *in vitro* mediante la adición de las ARN polimerasas apropiadas y los nucleótidos marcados apropiados. Las sondas de hibridación pueden marcarse con una variedad de grupos informadores, por ejemplo, radionúclidos tales como ^{32}P o ^{35}S , o marcadores enzimáticos, tales como fosfatasa alcalina acoplada a la sonda mediante sistemas de acoplamiento avidina/biotina, y semejantes. Los polinucleótidos pueden usarse en análisis Southern o Northern, hibridación sobre mancha, u otras tecnologías basadas en membrana; en tecnologías de PCR; o en tira reactiva, pin, ensayos ELISA, o micromatrices que utilizan fluidos o tejidos de biopsias del sujeto para detectar la presencia o niveles de un microbio. Dichos métodos cualitativos o cuantitativos son muy conocidos en la técnica.

En un aspecto particular, las secuencias de ácido nucleico pueden ser útiles en varios ensayos marcadas por métodos estándar, y añadidas a una muestra de fluido o tejido de un sujeto en condiciones adecuadas para hibridación y/o amplificación. Después de un periodo de incubación adecuado, la muestra se lava y la señal se cuantifica y compara con un valor estándar. Si la cantidad de señal en la muestra de ensayo está significativamente alterada respecto a la de una muestra control comparable, la presencia de niveles alterados de secuencias de nucleótidos en la muestra indica la presencia o niveles del microbio. Dichos ensayos también pueden usarse para evaluar la eficacia de un régimen de tratamiento particular en estudios de animales, en estudios clínicos, o en la monitorización del tratamiento de un sujeto.

Con el fin de proporcionar la base para el diagnóstico de la presencia o niveles de un microbio, se establece un perfil normal o estándar para expresión. Esto puede conseguirse combinando fluidos corporales o extractos celulares tomados de sujetos normales, con un polinucleótido o un fragmento de éste, en condiciones adecuadas para hibridación y/o amplificación. Los niveles estándar pueden cuantificarse comparando los valores obtenidos de sujetos normales con los de un experimento en el que se usa una cantidad conocida de un polinucleótido sustancialmente purificado. Los valores estándar obtenidos de muestras normales pueden compararse con valores

obtenidos de muestras de sujetos tratados para crecimiento microbiano. La desviación entre los valores estándar y del sujeto se usa para establecer la presencia o niveles del microbio.

Una vez el microbio se identifica y se inicia un protocolo de tratamiento, los ensayos de hibridación y/o amplificación pueden repetirse de una manera regular para evaluar si el nivel de expresión en el sujeto empieza a disminuir respecto a lo que se observa en el sujeto normal. Los resultados obtenidos de ensayos sucesivos pueden usarse para mostrar la eficacia del tratamiento durante un periodo que varía de varios días a meses.

Lo usos diagnósticos particulares para los oligonucleótidos diseñados a partir de las secuencias de ácido nucleico pueden implicar el uso de PCR. Algunos oligómeros pueden sintetizarse químicamente, generarse enzimáticamente, o producirse *in vitro*. Los oligómeros consistirán preferiblemente en dos secuencias de nucleótidos, una con orientación con sentido (5'→3') y otra con orientación antisentido (3'→5'), empleadas en condiciones optimizadas para la identificación de una secuencia de nucleótido o condición específica. Los mismos dos oligómeros, conjuntos anidados de oligómeros, o incluso un conjunto degenerado de oligómeros puede emplearse en condiciones menos astringentes para la detección y/o cuantificación de secuencias de ADN o ARN relacionadas de cerca.

Los métodos que también pueden usarse para cuantificar la expresión incluyen radiomarcaje o biotilación de nucleótidos, coamplificación de un ácido nucleico control, y curvas estándar en las que los resultados experimentales se interpolan (Melby, P. C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229-236). La velocidad de la cuantificación de múltiples muestras puede acelerarse operando el ensayo en un formato ELISA en el que el oligómero de interés se presenta en varias diluciones y una respuesta espectrofotométrica o colorimétrica proporciona una cuantificación rápida.

En realizaciones adicionales, los oligonucleótidos o fragmentos más largos derivados de cualquiera de los polinucleótidos descritos en la presente memoria pueden usarse como dianas en una micromatriz. La micromatriz puede usarse para monitorizar el nivel de expresión de grandes números de genes simultáneamente (para producir una imagen de transcrito), y para identificar variantes genéticas, mutaciones y polimorfismos. Esta información puede usarse para determinar la función génica, para entender la base genética de la enfermedad, para diagnosticar la enfermedad, y para desarrollar y monitorizar las actividades de agentes terapéuticos. En una realización, la micromatriz se prepara y se usa según métodos conocidos en la técnica tales como los descritos en la solicitud PCT WO 95/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) y Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619).

En un aspecto, los oligonucleótidos pueden sintetizarse en la superficie de la micromatriz usando un procedimiento de acoplamiento químico y un aparato de aplicación de chorro de tinta, tal como el descrito en la solicitud PCT WO 95/251116 (Baldeschweiler et al.). En otro aspecto, puede usarse una matriz "con cuadrícula" análoga a una transferencia en mancha o rejilla (aparato HYBRIDOT, Life Technologies) para organizar y unir fragmentos de ADNc u oligonucleótidos a la superficie de un sustrato usando un sistema de vacío, procedimientos de unión térmica, UV, mecánica o química. En otro aspecto más, puede producirse una matriz a mano o usando dispositivos, materiales, y máquinas disponibles (incluyendo pipeteadores multicanal o instrumentos robóticos; Brinkmann, Westbury, N.Y.) y pueden incluir, por ejemplo, 24, 48, 96, 384, 1.024, 1.536, ó 6.144 sitios o pocillos (por ejemplo, como una placa con múltiples pocillos), o más, o cualquier otro múltiplo de 2 a 1.000.000 que proporciona en sí mismo el uso eficiente de instrumentación disponible comercialmente.

Con el fin de llevar a cabo el análisis de muestras usando las micromatrices, los polinucleótidos se extraen de una muestra biológica. Las muestras biológicas pueden obtenerse de cualquier fluido corporal (sangre, orina, saliva, flema, jugos gástricos, etc.), células cultivadas, biopsias, u otras preparaciones de tejidos. Para producir sondas, los polinucleótidos extraídos de la muestra se usan para producir secuencias de ácido nucleico que son complementarias a los ácidos nucleicos en la micromatriz. Si la micromatriz consiste en ADNc, los ARN antisentido son sondas apropiadas. Por lo tanto, en un aspecto, se usa ARNm para producir ADNc que, a su vez, y en presencia de nucleótidos fluorescentes, se usa para producir fragmentos o sondas de ARN antisentido. Estas sondas marcadas fluorescentemente se incuban con la micromatriz de manera que las secuencias de la sonda hibridan con los oligonucleótidos de ADNc de la micromatriz. En otro aspecto, las secuencias de ácido nucleico usadas como sondas pueden incluir polinucleótidos, fragmentos, y secuencias complementarias o antisentido producidas usando enzimas de restricción, tecnologías de PCR, y kits de oligomarcaje (Amersham Pharmacia Biotech) muy conocidos en el área de la tecnología de la hibridación.

En otra realización de la descripción, los polipéptidos de la invención o fragmentos funcionales o inmunogénicos u oligopéptidos de éstos, pueden usarse para cribar bibliotecas de compuestos en cualquiera de una variedad de técnicas de cribado de fármacos. El fragmento empleado en dicho cribado puede estar libre en disolución, fijado a un soporte sólido, localizado en una superficie celular, o localizado intracelularmente. La formación de complejos de unión, entre el polipéptido y el agente que se está ensayando, puede medirse.

Una técnica para el cribado de fármacos que puede usarse proporciona un cribado de alto rendimiento de compuestos que tienen una afinidad de unión adecuada para el polipéptido de interés como se describe en la solicitud publicada PCT WO 84/03564. En este método, se sintetizan grandes números de diferentes compuestos de ensayo pequeños en un sustrato sólido, tal como pins de plástico o alguna otra superficie. Los compuestos de

ensayo se hacen reaccionar con el polipéptido, o fragmentos de éste, y se lava. El polipéptido unido se detecta por métodos muy conocidos en la técnica. El polipéptido purificado también puede utilizarse para recubrir directamente placas para uso en las técnicas de cribado de fármacos mencionadas anteriormente. Alternativamente, pueden usarse anticuerpos no neutralizantes para capturar el polipéptido e inmovilizarlo en un soporte sólido.

- 5 En otra técnica, se pueden usar ensayos de cribado de fármacos competitivos en los que anticuerpos neutralizantes capaces de unirse al polipéptido específicamente compiten con un compuesto de ensayo para la unión al polipéptido. De esta manera, los anticuerpos pueden usarse para detectar la presencia de un compuesto de ensayo que comparte uno o más sitios de unión a antígeno con el anticuerpo.

Ejemplos

- 10 Los ejemplos descritos en la presente memoria son para el propósito de ilustrar realizaciones de la invención. Otras realizaciones, métodos, y tipos de análisis están en el alcance de personas expertas en las técnicas de diagnóstico molecular y no es necesario describirlas con detalle en adelante. Otras realizaciones en el alcance de la técnica se considera que son parte de esta invención.

Ejemplo 1: estimación del tamaño del genoma

- 15 Se creció la cepa M1^T de *Methanobrevibacter ruminantium* (DSM1093) en medio BY+ (medio basal, Joblin et al., 1990) que consiste en [g/l] NaCl (1), KH₂PO₄ (0,5), (NH₄)₂SO₄ (0,25), CaCl₂·2H₂O (0,13), MgSO₄·7H₂O (0,2), K₂HPO₄ (1), fluido de rumen clarificado (300 ml), dH₂O (360 ml), NaHCO₃ (5), resazurina (0,2 ml), L-cisteína-HCl (0,5), extracto de levadura (2), y disolución de elementos traza de Balch (10 ml) (elementos traza añadidos; Balch et al., 1979) que consiste en (g/l) ácido nitrilotriacético (1,5), MgSO₄·7H₂O (3), MnSO₄·H₂O (0,5), NaCl (1), FeSO₄·7H₂O (0,1), CoCl₂·6H₂O (0,1), CaCl₂ (0,1), ZnSO₄·7H₂O (0,1), CuSO₄·5H₂O (0,01), AlK(SO₄)₂·12H₂O (0,01), H₃BO₃ (0,01), Na₂MoO₄·2H₂O (0,01), NiSO₄·6H₂O (0,03), Na₂SeO₃ (0,02), y Na₂WO₄·2H₂O (0,02). El ADN genómico se extrajo congelando sedimentos celulares en N₂ líquido y molienda usando un mortero y mano pre-enfriados, esterilizados. Los homogenados celulares se incluyeron en tapones de agarosa y las manipulaciones posteriores se llevaron a cabo en los tapones para reducir la cizalla física del ADN genómico. Las digestiones se llevaron a cabo con endonucleasas de restricción y los fragmentos de ADN se separaron usando electroforesis en gel en campo pulsado (PFGE).

Ejemplo 2: clonación y secuenciación del ADN

- El ADN del genoma de *M. ruminantium* se secuenció por Agencourt Biosciences Corporation (Massachusetts, EEUU) usando una estrategia de clonación al azar aleatoria (Fleischmann et al., 1995) y por MacroGen Corporation (Rockville, MD, EEUU) usando pirosecuenciación. Brevemente, se construyeron bibliotecas de ADN de *M. ruminantium* en *Escherichia coli* por disrupción física aleatoria de ADN genómico y separación de los fragmentos por electroforesis en gel. Los fragmentos grandes en el intervalo de 40 Kb se recuperaron del gel y se usaron para generar una biblioteca de fósido de insertos grandes. Los fragmentos de ADN en el intervalo de 2 a 4 Kb se recuperaron y se usaron para generar una biblioteca de plásmidos de insertos pequeños. Los clones resultantes tanto de la biblioteca de insertos grandes como pequeños se crecieron, y su ADN de fósido o plásmido se recuperó y secuenció usando tecnología de secuenciación de alto rendimiento. Un número suficiente de clones se secuenció para proporcionar una cobertura teórica de 8 veces del genoma de *M. ruminantium*. La pirosecuenciación se llevó a cabo en fragmentos de ADN genómico cortados aleatoriamente para proporcionar una cobertura teórica final de 10 veces.

- 40 Ejemplo 3: Ensamblaje de secuencias y anotación de profago

- Las secuencias de ADN se alinearon para encontrar superposiciones de secuencia y se ensamblaron en secuencias contiguas (cóntigo) usando Paracel Genome Assembler (Paracel Inc, CA, EEUU) y el paquete Staden (Staden et al., 1998) en combinación con secuencia tanto de PCR estándar como inversa. Los cóntigos se analizaron usando el descubridor de marco de lectura abierto (ORF) GLIMMER (Gene Locator interpolated Markov Model ER, Delcher et al., 1999) y cada ORF se analizó por BLASTP con huecos (Basic Local Alignment Search Tool (Altschul et al., 1997) frente a las bases de datos del National Center for Biotechnology Information (NCBI) de nucleótido no redundante y proteínas. Los cóntigos de la secuencia de la fase preliminar de 8 veces se unieron aleatoriamente por unión artificial de secuencias para generar una "pseudomolécula" y se enviaron a The Institute for Genomic Research (TIGR, DC, EEUU) para autoanotación. Los cóntigos ensamblados de la pirosecuenciación de 10 veces se reanalizaron usando GLIMMER y los ORF se autoanotaron usando GAMOLA (Global Annotation of Multiplexed On-site Blasted DNA sequences; Altermann y Klaenhammer, 2003). Los ORF se clasificaron por función usando las agrupaciones de la base de datos de proteínas ortólogas (COG) (umbral 1e-02) (protocolo de transferencia de hipertexto://world wide web.pnas.org/cgi/content/full/102/11/3906; Tatusov et al., 2001).

- Los restos de proteína se determinaron por HMMER (protocolo de transferencia de hipertexto://hmmer.wustl.edu) usando las bibliotecas PFAM HMM y TIGRFAM, con alineamiento global y local (protocolo de transferencia de hipertexto://pfam.wustl.edu) y modelos TIGRFAM HMMs estándar y de modo fragmento (protocolo de transferencia de hipertexto://world wide web.tigr.org/TIGRFAMs) respectivamente (umbral 1e-02). Los ARNt se identificaron usando TARNSCAN-SE (Lowe y Eddy, 1997) y las repeticiones de nucleótidos se identificaron usando el paquete de

software KODON (Applied Maths, Austin, TX, EEUU) y REPUTER (Kurtz y Schleiermacher, 1999). Las anotaciones automatizadas se verificaron posteriormente manualmente. Se construyeron visualizaciones de atlas de genoma usando GENEWIZ (Jensen et al., 1999) y se generaron estructuras de datos subyacentes por algoritmos personalizados de desarrollo propio. Las reconstrucciones de ruta a partir del ORFoma predicho de *M. ruminantium* se llevaron a cabo en conjunción con la base de datos en línea KEGG (Kyoto Encyclopedia of Genes and Genomes, Kanehisa et al., 2004) usando software de desarrollo propio (PathwayVoyager; Altermann y Klaenhammer, 2005).

Ejemplo 4: resultados de la secuenciación y análisis

La estimación del tamaño del genoma de *M. ruminantium* por digestión con enzimas de restricción de ADN genómico y dimensionado de los fragmentos mediante PFGE, indicó un único cromosoma de aproximadamente 2,5-2,9 Mb. La secuenciación inicial de clones con inserto grande y pequeño (cobertura preliminar de 6 veces) y ensamblaje de la secuencia en cóntigos indicó que una región de 40 Kb del genoma estaba altamente sobre-representada (>20 veces), particularmente en la biblioteca de insertos pequeños. Esto se debió posiblemente a un plásmido de alto número de copias (aunque no se han identificado ADN extracromosómicos) o un bacteriófago lisogénico que se había replicado durante el crecimiento del cultivo usado para la extracción de ADN. Debido a este sesgo de secuencia grande, se llevó a cabo una secuenciación adicional (cobertura de genoma teórica de 2 veces) sólo para los clones con insertos pequeños rindiendo una cobertura final de 8 veces a partir de secuenciación de Sanger. La secuencia de la fase preliminar de 8 veces se ensambló en 756 cóntigos que se unieron mediante 105 soportes. La pirosecuenciación adicional se llevó a cabo hasta una cobertura adicional -10 veces y la incorporación de estas secuencias en el ensamblaje resultó en que el número de cóntigos cayó hasta 27. El cierre de huecos posterior usando técnicas de PCR inversa y de rango largo redujo el número de cóntigos hasta 14, permaneciendo un ensamblaje incorrecto.

Durante la fase de secuenciación de alto rendimiento, se observó un sesgo en la cobertura de la secuencia hacia una región (~50Kb) con un contenido de G+C significativamente mayor inmediatamente adyacente a una región con bajo G+C (~12Kb). El análisis de la secuencia del genoma mediante GAMOLA y GeneWiz dio lugar al descubrimiento de una región prominente con alto GC localizada inmediatamente adyacente a una gran adición con bajo GC. Los análisis detallados de la región con alto G+C revelaron la presencia de productos génicos con similitudes con una integrasa relacionada con fago, la subunidad grande de la terminasa de fago, una proteína portal de fago, una proteína de cápside de fago, y una peptidasa predicha que actúa como lisina de fago (FIG. 3). Estos productos génicos se usaron como puntos de anclaje para la estructura global del profago predicho de *M. ruminantium*, designado ϕ mr. Sobre la base de los análisis de las estructuras secundarias del ADN, se identificaron los sitios probables de integración del fago *attL* y *attR* (FIG. 1A). La integración del fago en el sitio *att* parece haber disrumpido una posible proteína de membrana codificada por los ORF 1980 y 2069, y este gen puede albergar el sitio de integración original para el genoma del fago ϕ mr, *attB*.

La estructura general (FIG. 1 B) y secuencia de ADN (FIG. 4A) de ϕ mr se determinaron sobre la base de la estructura modular reconocida comúnmente de genomas de fago combinada con similitudes con bases de datos de secuencia y funcionales. Véase, por ejemplo, Altermann E, Klein JR, Henrich B. Primary structure and features of the genome of the *Lactobacillus gasseri* temperate bacteriophage (ϕ)adh. Gene. 1999 ago 20;236(2):333-46; Desiere F, Lucchini S, Canchaya C, Ventura M, Brüssow H. Antonie Van Leeuwenhoek. Comparative genomics of phages and prophages in lactic acid bacteria. 2002 ago; 82(1- 4):73-91. El ORFoma predicho del fago ϕ mr se clasificó con éxito en módulos que codifican la integración del fago, replicación y encapsidación del ADN, proteínas estructurales del fago, y un casete de lisis, y aproximadamente 40% de los ORF del fago se caracterizaron funcionalmente. Una estructura semejante a un terminator en una región no codificadora grande (244 pb), flanqueada por un gran número de repeticiones directas e indirectas y determinada en el módulo de replicación del ADN se caracterizó como un origen posible de la replicación del ADN. Varios genes en la secuencia del genoma del fago se predijeron en la cadena antisentido y éstas coincidieron con regiones bajas en GC. Debe determinarse si estos genes inactivan la función del fago o indican un ensamblaje incorrecto en el genoma del fago.

La región baja en GC entre la lisina del fago predicha y *attR*, se encontró que albergaba un sistema de modificación de azufre de ADN, *dnd* (degradación durante electroforesis), incluyendo una ADN metiltransferasa de adenina m6 de restricción de tipo II y un regulador transcripcional que probablemente sea específico para el sistema *dnd*. Además, se identificaron estructuras de ARN no codificadoras tanto en como flanqueando el genoma del fago. En el módulo de replicación de ADN predicho, se identificó un *rbcl*. *rbcl* representa un elemento estabilizador de ARN 5' UTR de *Chlamydomonas reinhardtii*. Se piensa que la familia está implicada en la estabilización del gen *rbcl* que codifica la subunidad grande de ribulosa-1,5-bisfosfato carboxilasa. Las mutaciones en esta familia pueden dar lugar a una aceleración de 50 veces en la degradación del transcrito.

Flanqueando el genoma del fago, se identificaron tres estructuras de intrón grupo I. Los intrones catalíticos del grupo I son ribozimas grandes de auto-corte y empalme. Catalizan su propia escisión de precursores de ARNm, ARNt y ARNr en un amplio rango de organismos. La estructura secundaria central consiste en 9 regiones apareadas (P1-P9). Éstas se pliegan en esencialmente dos dominios - el dominio P4-P6 (formado a partir del apilamiento de hélices P5, P4, P6 y P6a) y el dominio P3-P9 (formado a partir de las hélices P8, P3, P7 y P9). La marca de estructura secundaria para esta familia representa sólo esta parte central conservada. Los intrones catalíticos del grupo I tienen frecuentemente ORF largos insertados en regiones bucle. Estas estructuras de ARN no codificadoras están

localizadas en las regiones no codificadoras entre 5' de ORF 1980 (SEQ ID NO: 74), 3' de ORF 2065 (SEQ ID NO: 141) y *attR* y 5' de ORF 2069 (SEQ ID NO: 142).

Ejemplo 5A: genes de fago

El descubrimiento de una secuencia de profago en la secuencia del genoma de *M. ruminantium* fue inesperado. No ha habido ningún reporte previo de que la cepa M1 de *Methanobrevibacter ruminantium* (DSM 1093) sea susceptible bien a fago lítico o lisogénico, aunque ha habido reportes de fago siendo identificado para otras especies de *Methanobrevibacter* (Baresi y Bertani, 1984; Knox y Harris, 1986). La secuencia del profago ϕ mrú es significativamente mayor en contenido G+C que el genoma de *M. ruminantium* circundante lo que sugiere que se originó de otro organismo. Los niveles observados de homología no sugieren un huésped obvio del que se originó, e indican que ϕ mrú es distinto de cualquier otro fago encontrado hasta la fecha.

La secuencia de ADN de ϕ mrú está insertada en una proteína de membrana predicha de *M. ruminantium* y está flanqueada por secuencias de ADN con estructuras secundarias consistentes con sitios *attL* y *attR*. A pesar de la ausencia de una homología fuerte con otras proteínas conocidas, todas las características de módulo funcional de un fago podrían identificarse en la secuencia de ϕ mrú. Una característica interesante de la secuencia es una región con bajo G+C en el extremo 3' que muestra homología con proteínas implicadas en un sistema de modificación de azufre de ADN (*dnd*). Estos genes están localizados en 5' del sitio de unión *attR* de ϕ mrú y probablemente por lo tanto se llevaron en el genoma de *M. ruminantium* durante la integración del fago. La región codifica varios ORF asociados con *dnd* (*dnd* 1, 2, y 3), y una subunidad de metilasa de tipo II junto con un posible regulador transcripcional.

El fenotipo *dnd* sensibiliza su ADN a degradación durante electroforesis. Los análisis de las funciones respectivas de los ORF de *dnd* sugirieron una incorporación de azufre o una sustancia que contenía azufre en el genoma de los huéspedes. También se descubrió que el fenotipo *Dnd* existía en ADN de especies bacterianas extendidas de origen variable y hábitat diverso. Se encontraron agrupaciones génicas organizadas de forma similar en varios genomas bacterianos que representan diferentes géneros y en ADN de organismos marinos, lo que sugiere que dicha modificación es un fenómeno extendido. Se demostró que ocurría una coincidencia entre el fenotipo *Dnd* y la modificación de ADN por azufre en varios genomas bacterianos representativos por los experimentos de marcaje con (³⁵)S *in vivo* (Zhou X, He X, Liang J, Li A, Xu T, Kieser T, Helmann JD, Deng Z. A novel DNA modification by sulphur. *Mol Microbiol.* 2005 sep;57(5):1428-38).

Los sistemas R/M de tipo II son los más simples y los más prevalentes. En lugar de funcionar como un complejo, la metiltransferasa y endonucleasa están codificadas como dos proteínas separadas y actúan independientemente. No hay especificidad de proteína. Ambas proteínas reconocen el mismo sitio de reconocimiento, y por lo tanto compiten para actividad. La metiltransferasa actúa como un monómero, metilando el dúplex una cadena cada vez. La endonucleasa actúa como un homodímero, que facilita la escisión de ambas cadenas. La escisión ocurre en una posición definida cerca de o en la secuencia de reconocimiento. En este punto, no está claro cómo la funcionalidad predicha actúa junto con el sistema *dnd*. Aun así, está claro que el fago tiene utilidad como un vehículo de administración de genes. En particular, la región de conversión lisogénica puede usarse como el locus de reemplazo de genes.

Es probable que el sistema *dnd* se transportó a *M. ruminantium* por el fago. Como tal, el papel del sistema *dnd* de ϕ mrú en la protección o modificación de ADN de *M. ruminantium* o extraño no se conoce. Otra característica interesante de la secuencia de ϕ mrú es el número de ORF codificados en la cadena antisentido. Estos ORF se corresponden con regiones con bajo GC y tienen concordancias BLAST débiles con proteínas de una variedad de organismos. Esto podría sugerir que estos genes se han acumulado en el genoma de ϕ mrú desde su integración en *M. ruminantium*. No está claro si estos ORF representan una acumulación en curso de inserciones que pueden dar lugar eventualmente a la inactivación y domesticación del fago o si ϕ mrú es totalmente activo.

Un gen de ϕ mrú con interés particular para la mitigación de metano es ORF2058 localizado en el casete de lisis. ORF 2058 está anotado como una peptidasa y tiene una concordancia de Familia de Proteínas (Pfam) (Puntuación: -13,7, valor E: 0,00054) con las proteínas de la familia Peptidasa C39. Estas proteínas son cisteína peptidasas y son parte del clan mayor de peptidasas CA como se define por la base de datos de peptidasas MEROPS (Rawlings et al., 2006). La familia de peptidasas C39 está asociada habitualmente con transportadores ABC y funciona como proteasas de maduración durante la exportación y procesamiento de bacteriocinas. El clan de la peptidasa CA también incluye las cisteínas endopeptidasas virales tales como las endoisopeptidasas de fago arqueal C71 que escinden los péptidos de entrecruzamiento de la pared celulares de arqueal. Las paredes celulares de arquea metanogénica que pertenecen a la familia *Methanobacteriales* contienen cadenas paralelas de pseudomureína, un polímero de ácido N-acetil-L-talosaminurínico entrecruzado por un péptido. Las endoisopeptidasas pseudomureína C71 son capaces de escindir los entrecruzamientos del péptido de la pared celular de arquea y lisar las células.

Sobre la base de la localización y sintenia con la endoisopeptidasa pseudomureína del fago Ψ M2 de *Methanothermobacter marburgensis* (FIG. 3), ORF 2058 puede tener un papel como un gen de lisina de metanógeno que codifica la enzima lítica implicada en la lisis celular antes de la liberación de la progenie del fago. El alineamiento de ORF 2058 con PeiP de *M. marburgensis* y PeiW de *M. wolfeii* (FIG. 5) muestra una baja homología

global entre las proteínas. Sin embargo, existe conservación de los residuos de histidina y ácido aspártico implicados en la triada catalítica de endoisopectidasa y un residuo de cisteína en ORF 2058 está posicionado cerca de la cisteína conservada de PeiP y PeiW que forma el tercer sitio conservado de la triada catalítica (Makarova et al., 1999, Luo et al., 2002). Además, el resto Gly-His-Tyr que rodea el residuo de His catalítico en PeiP y PeiW también se encuentra en ORF 2058. Estas observaciones indican que ORF 2058 es un gen de lisina de ϕ mru que funciona para lisar células de *M. ruminantium* durante el ciclo lítico del fago. Las diferencias observadas entre ORF 2058 y PeiP y PeiW pueden reflejar diferentes entrecruzamientos de péptidos de la pared celular arqueal y por lo tanto especificidad de sustrato de la peptidasa.

Ejemplo 5B: inducción de fago

Se creció la cepa M1^T de *Methanobrevibacter ruminantium* (DSM1093) en medio BY+ (medio basal, Joblin et al., 1990) que consiste en [g/l] NaCl (1), KH₂PO₄ (0,5), (NH₄)₂SO₄ (0,25), CaCl₂·2H₂O (0,13), MgSO₄·7H₂O (0,2), K₂HPO₄ (1), fluido de rumen clarificado (300 ml), dH₂O (360 ml), NaHCO₃ (5), resazurina (0,2 ml), L-cisteína-HCl (0,5), extracto de levadura (2), y disolución de elementos traza de Balch (10 ml) (elementos traza añadidos; Balch et al., 1979) que consiste en (g/l) ácido nitrilotriacético (1,5), MgSO₄·7H₂O (3), MnSO₄·H₂O (0,5), NaCl (1), FeSO₄·7H₂O (0,1), CoCl₂·6H₂O (0,1), CaCl₂ (0,1), ZnSO₄·7H₂O (0,1), CuSO₄·5H₂O (0,01), AlK(SO₄)₂·12H₂O (0,01), H₃BO₃ (0,01), Na₂MoO₄·2H₂O (0,01), NiSO₄·6H₂O (0,03), Na₂SeO₃ (0,02), y Na₂WO₄·2H₂O (0,02).

A densidades ópticas (DO), medidas a una longitud de onda de 600 (DO₆₀₀) entre 0,10 y 0,14, *M. ruminantium* se pulsó con 1 ml y 2 ml de aire estéril (~160 a 320 µl de oxígeno), respectivamente (Figura 1C) y 2 µg/ml de MitomicinaC (Figura 1D). Pudieron observarse curvas de lisis típicas para ambos pulsos, con tiempos de latencia de ~90 min para pulso de aire. Los resultados iniciales para el pulso con MitomicinaC indican un periodo de latencia muy corto. Para verificar la escisión del fago del genoma del huésped, se diseñaron 2 oligonucleótidos, orientados hacia ambos sitios de unión del fago, respectivamente (R1F: caaagagagattaagaagcagacg; SEQ ID NO: 146 y L2R agtagtggtggaatcagtgaaaagg; SEQ ID NO: 147). Esta pareja de cebadores sólo produce un amplicón si el genoma del fago se recirculariza después de escisión.

La Figura 1E representa los experimentos de escisión iniciales cuando *M. ruminantium* se pulsó con aire. Después de la inducción, se encontró un amplicón claro y no ambiguo del tamaño esperado, indicando una escisión y recircularización exitosas. Una banda similar, aunque más débil, también se encontró en células de *M. ruminantium* no inducidas, indicando que ϕ mru tiene la capacidad de escindirse espontáneamente durante el crecimiento normal, no pulsado.

Ejemplo 5C: bioensayos de enzimas líticas

El polipéptido codificado por ORF 2058 es útil como una enzima lítica específica de metanógenos del rumen y se ha subclonado en un vector de expresión de *E. coli* para la producción de proteína recombinante. ORF 2058 se amplificó por PCR usando los cebadores Mbbrum11for22 (1122For, cac cat ggt tag att cag cag aga c; SEQ ID NO:148) y Mbbrum11rev22 (1122Rev, tca tgc agg aca gac aac ata gta g; SEQ ID NO: 149) en 150 µL de volumen de reacción que contenía: 121,5 ng M. de ADN genómico de la cepa M1 de *M. ruminantium*; 0,2 µM cebadores 1122For y 1122Rev; 15 µL de tampón Accuprime Pfx (con dNTP, InVitrogen); 2,4 µL de Accuprime Pfx (InVitrogen). Las condiciones de PCR fueron 95°C durante 2 min desnaturalización inicial seguido de 35 ciclos de 95°C durante 15 segundos, 55°C durante 30 segundos y 68°C durante 40 segundos. No se usó extensión final. El producto de PCR se purificó y cuantificó usando un Nanodrop (Thermo Scientific, GA, EEUU).

Clonación de ORF 2058: El ORF 2058 amplificado por PCR se clonó bien en los vectores Topo pET 100 o pET 151-D (InVitrogen) según las recomendaciones del fabricante, y se transformaron en células TOP 10 químicamente competentes (InVitrogen). Los transformantes se analizaron por PCR de colonias, y el ADN plasmídico se purificó y secuenció. Los clones con secuencias de ADN concordantes con las de ORF 2058 se seleccionaron.

Expresión de ORF 2058: El ADN plasmídico de clones que contenían insertos verificados de ORF 2058 se transformaron mediante electroporación en células BL21* o Rosetta 2 electro-competentes. Se encontró que las mejores condiciones de crecimiento para la expresión de proteína ORF 2058 soluble eran en medio LB, con la inducción llevada a cabo entre 0,48-0,6 de absorbancia 600 nm usando 0,5 mM IPTG y continuando el crecimiento durante aproximadamente seis horas a 30°C. Las células se recogieron por centrifugación y se congelaron a -20°C.

Lisis celular: El sedimento celular se descongeló y se resuspendió en el tampón siguiente (pH 7,5): 300 mM NaCl, 2 mM DTT, 10 mM imidazol, 20 mM Tris, 20% glicerol, 1% Tritón-X, 5 mM CaCl₂, y 10 mM MgCl₂. Se añadió lisozima a 1 mg/ml de concentración final, seguido de incubación en hielo con agitación suave durante 30 min. Se añadieron ADNasa I y ARNasa I cada una a 5 µg/ml de concentración final seguido de incubación en hielo con agitación suave durante 30 min. El lisado celular se centrifugó a 12.000 rpm durante 15 min y el lisado crudo se filtró a través de un filtro de 0,8 µm.

Cromatografía de afinidad de níquel: El sobrenadante filtrado del procedimiento de lisis celular se aplicó a una columna de afinidad de níquel de 80 mL y se eluyó usando un gradiente de 20 mM a 250 mM imidazol en el tampón siguiente (pH 7,5): 300 mM NaCl, 2 mM DTT, 20 mM Tris, y 20% glicerol. Las fracciones eluidas de la columna que contenían la proteína ORF 2058 expresada se concentraron usando una célula de ultra filtración Millipore con una

membrana con un punto de corte de peso molecular de 10.000 kDa. La construcción ORF 2058 en pET100 expresada en células de *E. coli* BL21* se eluyó de la columna de níquel con el tampón de elución siguiente, pH 8,2 (20 mM Tris, 250 mM imidazol, 300 mM NaCl, 10 mM b-mercaptoetanol, 10% glicerol), y la enzima se almacenó en un tampón al que se añadieron glicerol adicional y ditiotretitol para conseguir una concentración final de 40% glicerol, 1 mM ditiotretitol, pH 8,2).

Desalado: El desalado de la proteína concentrada expresada a partir de la construcción pET 151 en células Rosetta 2 se llevó a cabo usando una columna de 250 mL BioGel P6 DG (BioRad, CA, EEUU) con el tampón siguiente (pH 7,0): 20 mM MOPS, 1 mM DTT, 300 mM NaCl, y 20% glicerol. Las fracciones de la columna se concentraron como se ha descrito anteriormente y la muestra final se filtró y se congeló rápidamente en nitrógeno líquido antes de almacenarla a -20°C.

Lisis de células de *M. ruminantium* en reposo: Se crecieron cultivos de cinco ml de *M. ruminantium* M1 (DSM 1093) en medio BY+ en tubos Hungate hasta fase logarítmica tardía y las células se recogieron por centrifugación de los tubos Hungate a 5.000 x g a temperatura ambiente durante 30 minutos. Los tubos se pusieron en una cámara anaeróbica (atmósfera 95% CO₂- 5% H₂, Coy Laboratory Products, MI, EEUU) donde el sobrenadante se desechó y las células de 10 ml de cultivo se resuspendieron en 1 ml de tampón MOPS pH 6,8 (50 mM MOPS, 5 mM CaCl₂, 1 mM ditiotretitol). La suspensión celular se ajustó a una DO (600 nm) de -0,12 por dilución con tampón MOPS adicional.

La suspensión celular estandarizada (50 µl) se dispensó en una placa de microtitulación y se añadieron concentraciones variadas de enzima lítica ORF 2058 (preparada a partir de la construcción pET 100) y el volumen total de la reacción se llevó a 250 µl con tampón. Las mezclas de célula y proteína se incubaron a 37°C y se registraron las lecturas de DO. Los efectos de las adiciones de enzima (µg de enzima añadida por ensayo) en células de *M. ruminantium* en reposo se muestran en la FIG. 7. Las adiciones de enzima disminuyeron las lecturas de DO a 600 nm de las células suspendidas de una manera dependiente de la dosis comparado con las células control sin enzima añadida. Esto indica que la enzima lítica ORF 2058 es capaz de atacar y lisar células de *M. ruminantium* en reposo en condiciones anaeróbicas.

Lisis de células de *M. ruminantium* en crecimiento: Se creció *M. ruminantium* en medio RM02. El medio RM02 estaba compuesto por los ingredientes siguientes (g/L): KH₂PO₄ (1,4), (NH₄)₂SO₄ (0,6), KCl (1,5), disolución de elementos traza SL10 (1 ml), disolución de selenito/tungstato, (1 ml), 0,1% (p/v) disolución de resazurina (4 gotas). Los componentes se mezclaron e hirvieron bajo 100% CO₂ sin O₂ y se enfriaron en un baño de hielo mientras se burbujeara 100% CO₂. Después de enfriar, se añadieron NaHCO₃ (4,2 g) y L-cisteína-HCl.H₂O (0,5 g) y 9,5 ml del medio se dispensaron en tubos Hungate mientras se desgaseaban los tubos con 100% CO₂. Los tubos se autoclavaron y se almacenaron en oscuridad durante 24 h antes del uso. Antes de la inoculación, se añadió NoSubRFV (0,5 ml por tubo, que contenía sustratos, extracto de levadura, vitaminas). Después de la inoculación, los tubos se gasearon con 80% CO₂/20% H₂ a 25 lb/in². Se creció *M. ruminantium* hasta mid-log (DO 600 nm ~0,1) punto en el que la enzima lítica ORF 2058 (preparada a partir del clon pET 151 D Topo) se añadió a los cultivos a concentraciones variadas. La incubación de los cultivos continuó y las lecturas de DO se registraron. El efecto de las adiciones de enzima en el crecimiento de *M. ruminantium* y formación de metano (% de producción de metano respecto al control sin adición después de 217 horas de crecimiento se indica entre paréntesis) se muestran en la FIG. 8. Los resultados muestran que la enzima lítica ORF 2058 afectaba dramáticamente el crecimiento de *M. ruminantium* de una manera dependiente de la dosis, disminuyendo la DO 600 nm de los cultivos en crecimiento en 2 horas de la adición. Los dos niveles más altos de adición de enzima también redujeron la formación de metano hasta un grado similar al de la adición de cloroformo (adición 100 µl /10 ml de cultivo).

Ejemplo 6: Visión global

Un descubrimiento inesperado de la secuenciación del genoma de *M. ruminantium* fue la presencia de una secuencia de profago. El análisis de la secuencia del genoma identificó una región con un contenido en GC inusualmente alto que contenía un número de genes relacionados con fago. La estructura global del profago predicho se identificó por análisis bioinformáticos adicionales y se designó como φmru. Aproximadamente el 40% de los genes del fago se asignaron a grupos funcionales discretos incluyendo integración del fago, replicación y encapsidación del ADN, proteínas estructurales del fago y lisis. Se encontró que las secuencias de ADN que flanquean el genoma del fago representaban sitios potenciales para la integración del fago (*attL* y *attR*). El fago parece haberse insertado a sí mismo en una posible proteína de membrana de *M. ruminantium* que probablemente alberga el sitio de integración original del metanógeno para el genoma del fago φmru, *attB*. Además, se piensa que una estructura semejante a terminador encontrada en el módulo de replicación del ADN representa un origen de replicación del ADN del fago. Una región con bajo GC en el extremo 3' del genoma del fago alberga lo que parece ser un sistema de modificación del ADN por azufre, incluyendo un gen que probablemente controla la expresión del sistema dnd. Estos genes probablemente se llevaron al genoma de *M. ruminantium* durante la integración del fago y su papel respecto a modificar el ADN del fago, huésped o extraño permanece sin elucidar. La retención del sistema dnd por *M. ruminantium* sugiere que ha impartido un beneficio al huésped. Sin embargo, el papel del sistema dnd de φmru en la modificación de ADN de *M. ruminantium* o extraño todavía está bajo investigación.

Otra característica interesante de la secuencia de ϕ mrú es el número de genes codificados en la cadena antisentido que se corresponde con regiones bajas en GC y tienen concordancias débiles con proteínas de una variedad de organismos. Esto sugiere que estos genes se han acumulado en el genoma de ϕ mrú desde su integración en *M. ruminantium* y puede ser que estos genes representen una acumulación en curso de inserciones que podrían dar lugar eventualmente a la inactivación del fago y domesticación del fago. El alto contenido en GC de la secuencia del fago ϕ mrú comparado con el genoma de *M. ruminantium* sugiere que se originó de otro organismo. Sin embargo, el huésped previo no es obvio ya que las proteínas de ϕ mrú parecen de alguna forma únicas por comparación con otros fagos encontrados hasta la fecha.

Los genes de ϕ mrú de interés notable respecto a la mitigación de metano son aquellos localizados en el casete de lisis. Un gen en particular codifica una proteína con similitud con la familia de peptidasas C39. Esta familia de peptidasas incluye, entre otras, cisteína endopeptidasas virales tales como las endoisopeptidasas de fago arqueal C71 que escinden los péptidos de entrecruzamiento de pseudomureína que compone las paredes celulares de *Methanobrevibacter*. Sobre la base de la localización génica en el genoma del fago y sintenia con endoisopeptidasas pseudomureína de otros genomas de fago de metanógenos no de rumen, este gen puede tener un papel como un gen de lisina que codifica la enzima lítica implicada en la lisis celular antes de la liberación de la progenie del fago. Este gen y su enzima codificada tienen un interés obvio como posible mecanismo de control para *M. ruminantium* y otros metanógenos del rumen con paredes celulares similares.

El fago de rumiantes y sus enzimas que están implicadas en lisar células huésped representan oportunidades significativas para controlar las poblaciones de metanógenos y otros miembros de la comunidad (bacterias, protozoos y hongos) en el rumen. Además, es posible identificar dianas enzimáticas del huésped clave que son susceptibles de inhibición por proteínas de fago a través de la comprensión de los ciclos de vida del fago. Los inventores han sondeado la composición de fago de rumen en vacas, ovejas y ciervo y han mostrado que presentan una variación temporal en los números y tipo. También se han identificado los aislados de metanógeno de Nueva Zelanda que están afectados por fago. Se han usado cultivos puros de metanógenos para evaluar las enzimas líticas de fago, y se han desarrollado técnicas basadas en cultivo y basadas en PCR para cribar para nuevos fagos. Se ha mostrado que los fagos purificados de muestras de rumen son factibles de análisis de secuencia de ADN aleatoria lo que permite descubrir enzimas del fago.

Hay varias ventajas para el uso de fago o sus enzimas en técnicas de mitigación para disminuir las emisiones de metano. Los fagos son miembros naturales de la comunidad microbiana del rumen y, así, no se verían como un tratamiento con antibiótico (y podría superar más fácilmente cualquier restricción reguladora). Los fagos son habitualmente específicos para un rango estrecho de huéspedes lo que permite potencialmente el direccionamiento seleccionado de metanógenos. La terapia con fagos se reconoce ahora como un tratamiento para organismos resistentes a antibióticos y generalmente se considera seguro. Una vez producidos, los fagos son habitualmente relativamente estables. La introducción de cepas de metanógenos en el rumen que son susceptibles a fagos podría tener efectos beneficiosos a largo plazo, particularmente si la inoculación ocurre a una edad temprana (por ejemplo, en corderos y terneros jóvenes). Se sabe que determinados metanógenos contienen genomas de fagos, son susceptibles a fagos líticos, o experimentan autólisis (sugere de enzimas líticas) incluyendo *Methanobrevibacter smithii* (cepa PS), *Methanobacterium bryantii* y *Methanobrevibacter* cepa MF-1. Un ejemplo notable de fago que se está usando para inhibir organismos problemáticos para la agricultura es el uso de fagos para tomar como diana *Escherichia coli*. 0157:H7.

Methanobrevibacter ruminantium se eligió para secuenciación de genoma debido a su prevalencia en el rumen bajo una variedad de condiciones dietéticas (sobre la base de datos de detección de cultivo y moleculares), la disponibilidad de cultivos, su factibilidad para crecimiento rutinario en el laboratorio, y la cantidad relativamente grande de estudios previos y antecedentes en la bibliografía disponibles para este organismo. La presente invención proporciona datos importantes respecto al genoma de *M. ruminantium*, y construye una imagen detallada del fago en el rumen. La secuencia del profago ϕ mrú proporciona reactivos específicos para la inhibición de *M. ruminantium* y para manipulaciones genéticas futuras para ayudar en la determinación de la función génica. El fago puede usarse para bloquear funciones/componentes conservados entre metanógenos para prevenir o reducir la formación de metano en el rumen.

REFERENCIAS

- Altermann E, Klaenhammer TR (2005) PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics* 6:60-66.
- 5 Altermann, E Klaenhammer TR (2003) GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. *OMICS: A journal of integrative biology* 7, 161-169.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389-3402.
- Balch WE, Fox GE, Magrum LJ, Woese CR, Wolfe RS (1979) Methanogens: reevaluation of a unique biological group. *Microbiological Reviews* 43, 260-296.
- 10 Baresi, L y Bertani, G. 1984. Isolation of a bacteriophage for a methanogenic bacterium. En *Abstracts of the Annual Meeting of the American Society for Microbiology*. Washington DC: American Society for Microbiology, p. 133.
- Bickle, T. A. y D. H. Kruger. 1993. Biology of DNA restriction. *Microbiol. Rev.* 57:434-450.
- Bult CJ, et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273, 1058-1073.
- 15 Coutinho PM, Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. En 'Recent Advances in Carbohydrate Bioengineering' (Eds HJ Gilbert, G Davies, B Henrissat y B Svensson) p. 3-12 (The Royal Society of Chemistry, Cambridge) (Carbohydrate Active Enzymes database, <http://www.cazy.org/>).
- Delcher AL, Harmon D1 Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Research* 27, 4636-4641.
- 20 Fleischmann, RD, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496-512.
- Fricke WF, Seedorf H, Henne A, Kruer M, Liesegang H, Hedderich R, Gottschalk G, Thauer RK (2006) The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H₂ for methane formation and ATP synthesis. *Journal of Bacteriology* 188, 642-658.
- 25 Godde JS, Bickerton A (2006) The repetitive DNAe called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *Journal of Molecular Evolution* 62, 718-729.
- Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR- associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Computational Biology* 1:474-483
- 30 Jansen R, Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* 43, 1565-1575.
- Jansen R, van Embden JD, Gaastra W, Schouls LM (2002) Identification of a novel family of sequence repeats among prokaryotes. *OMICS: A journal of integrative biology* 6, 23-33.
- Jensen LJ, Friis C, Ussery DW (1999) Three views of microbial genomes. *Research in Microbiology* 150, 773-777.
- 35 Joblin KN, Naylor GE, Williams AG (1990) Effect of *Methanobrevibacter smithii* on xylanolytic activity of anaerobic ruminal fungi. *Applied and Environmental Microbiology* 56, 2287-2295.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Research* 32, D277-D280.
- Kiener, A., Konig, H., Winter, J. y Leisinger, T. 1987. Purification and use of *Methanobacterium wolfeii* pseudomurein endopeptidase for lysis of *Methanobacterium thermoautotrophicum*. *J. Bacteriol.* 169, 1010-1016.
- 40 Knox, M. R. y Harris, J. E. 1986. Isolation and characterisation of a bacteriophage of *Methanobrevibacter smithii*. En *Abstracts of the XIV International Congress on Microbiology*. Manchester. International Union of Microbiological Societies.
- Kurtz S, Schleiermacher C (1999) REPuter fast computation of maximal repeats in complete genomes. *Bioinformatics* 15, 426-427.
- 45 Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25, 955-964.
- Loenen, W. y N. Murray. 1986. Modification enhancement by restriction alleviation protein (Ra1) of bacteriophage lambda. *J. Mol. Biol.* 190:11-22.

- Lucchini, S., F. Desiere, y H. Brussow. 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J. Virol.* 73:8647-8656.
- 5 Luo, Y. N., Pfister, P., Leisinger, T. y Wasserfallen, A. 2002. Pseudomurein endoisopeptidases PeiW and PeiP, two moderately related members of a novel family of proteases produced in *Methanothermobacter* strains. *FEMS Microbiol. Lett.* 208, 47- 51.
- Makarova, K. S., Aravind, L. y Koonin, E. V. 1999. A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. *Protein ScL* 8, 1714-1719.
- 10 Makarova KS, Grishin NV, Shabalina, SA, Wolf YI, Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* 1:7-32.
- Estadísticas de Nueva Zelanda 2005 (www.stats.govt.nz)
- Inventario de Gases de Efecto Invernadero de Nueva Zelanda 1990 - 2004. The National Inventory Report and Common Reporting Format. (2006) Ministry for the Environment. <http://www.mfe.govt.nz/publications/climate/nir-apr06/nir-apr06.pdf>.
- 15 Reeve JN, Nolling J Morgan RM, Smith DR (1997) Methanogenesis: genes, genomes and who's on first? *Journal of Bacteriology* 179, 5975-5986.
- Rawlings, N. D., Morton, F. R. y Barrett, A. J. 2006. MEROPS: the peptidase database. *Nucleic Acids Res.* 34, D270-D272.
- 20 Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, Latreille P, Kim K, Wilson RK, Gordon JI (2007) Genomic adaptations of *Methanobrevibacter smithii* to the human gut. *Proceedings of the National Academy of Sciences USA* 104, 10643-10648.
- Smith DR, et al. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* DH: Functional analysis and comparative genomics. *Journal of Bacteriology* 179, 7135-7155.
- 25 Smith PH, Hungate RE (1958) Isolation and characterization of *Methanobacterium ruminantium* n. sp. *Journal of Bacteriology* 75, 713-718.
- Staden R, Beal KF, Bonfield JK (1998) The Staden Package. *Methods in Molecular Biology: Bioinformatics Methods and Protocols* 132, 115-130.
- 30 Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT1 Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes *Nucleic Acids Research* 29, 22-28.

REIVINDICACIONES

1. Un polipéptido aislado: a) que comprende SEQ ID NO: 72; b) que es una variante funcional de SEQ ID NO: 72, y que comparte al menos 90% de identidad con la secuencia de longitud completa de SEQ ID NO: 72; o c) que es una variante funcional de SEQ ID NO: 72, y que comparte al menos 80% de identidad con la secuencia de longitud completa de SEQ ID NO: 72; en el que dichas variantes funcionales son capaces de inhibir una célula metanógena,
2. Un polinucleótido aislado que comprende: a) secuencia codificadora para la secuencia de longitud completa de SEQ ID NO: 72; b) secuencia codificadora para un polipéptido que comparte al menos 90% de identidad con la secuencia de longitud completa de SEQ ID NO: 72 y que es una variante funcional de SEQ ID NO: 72; c) secuencia codificadora para un polipéptido que comparte al menos 80% de identidad con la secuencia de longitud completa de SEQ ID NO: 72, y que es una variante funcional de SEQ ID NO: 72; o d) una secuencia complementaria a una cualquiera de (a) a (c), en el que dichas variantes funcionales son capaces de inhibir una célula metanógena.
3. Un vector a) que codifica el polipéptido de la reivindicación 1; o b) que comprende el polinucleótido de la reivindicación 2.
4. Una célula huésped: a) que está modificada genéticamente para codificar el polipéptido de la reivindicación 1; b) que está modificada genéticamente para comprender el polinucleótido de la reivindicación 2; o c) que comprende el vector de la reivindicación 3.
5. La célula huésped de la reivindicación 4, que es una célula procariota.
6. La célula huésped de la reivindicación 5, que es *Escherichia coli*.
7. Una molécula conjugada o molécula de fusión que comprende el polipéptido de la reivindicación 1.
8. Una molécula conjugada o molécula de fusión que comprende el polipéptido de la reivindicación 1, y que comprende además un compuesto anti-metanogénesis, una secuencia señal, un anticuerpo o fragmento de anticuerpo, un ácido nucleico peptídico, un péptido antimicrobiano, o un antibiótico.
9. Un fago ϕ mrú aislado preparado por ingeniería para comprender al menos un polinucleótido de la reivindicación 2.
10. Un método para inhibir una célula metanógena que comprende poner en contacto la célula con el polipéptido de la reivindicación 1, poner en contacto la célula con la molécula conjugada o molécula de fusión de la reivindicación 7 o reivindicación 8, o poner en contacto la célula con el fago de la reivindicación 9.
11. El método de una cualquiera de la reivindicación 10, en el que la célula es un metanógeno del rumen.
12. El método de la reivindicación 11, en el que la célula es *Methanobrevibacter ruminantium*.
13. El método de la reivindicación 12, en el que la célula es la cepa M1^T de *Methanobrevibacter ruminantium* (DSM1093).

FIG. 1

(A)ATAATT AATTATT

A

B

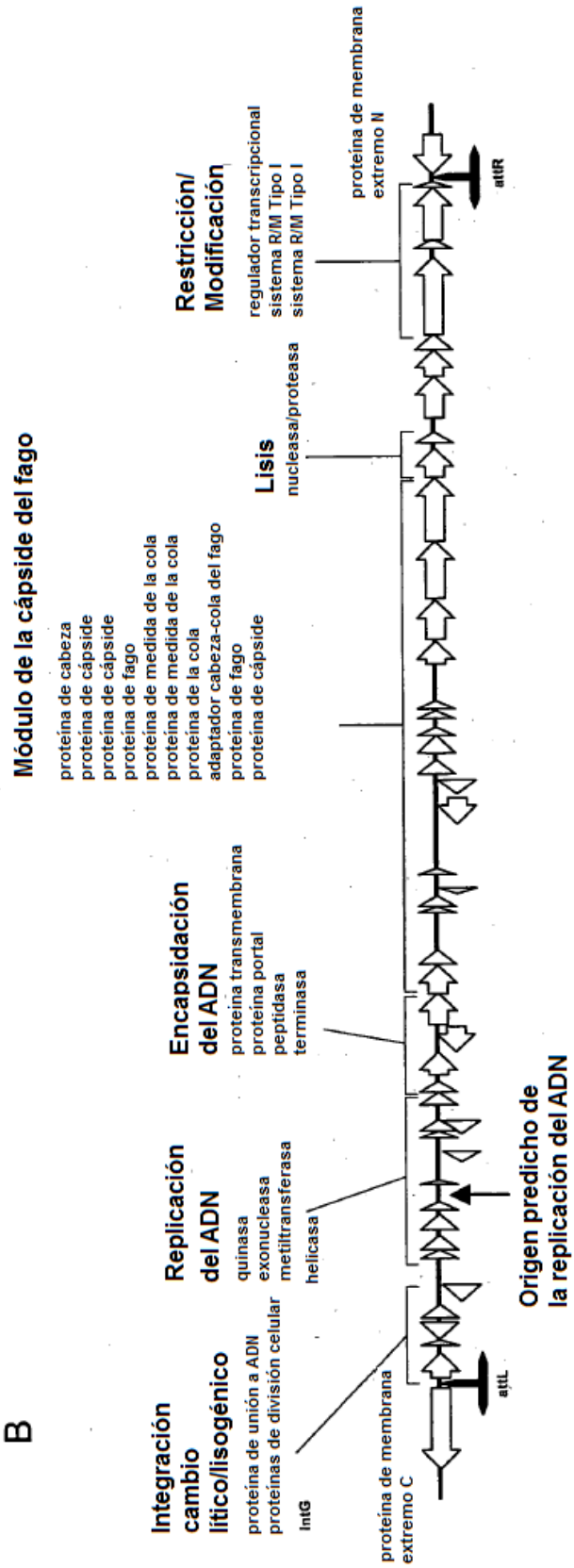


FIG. 1C

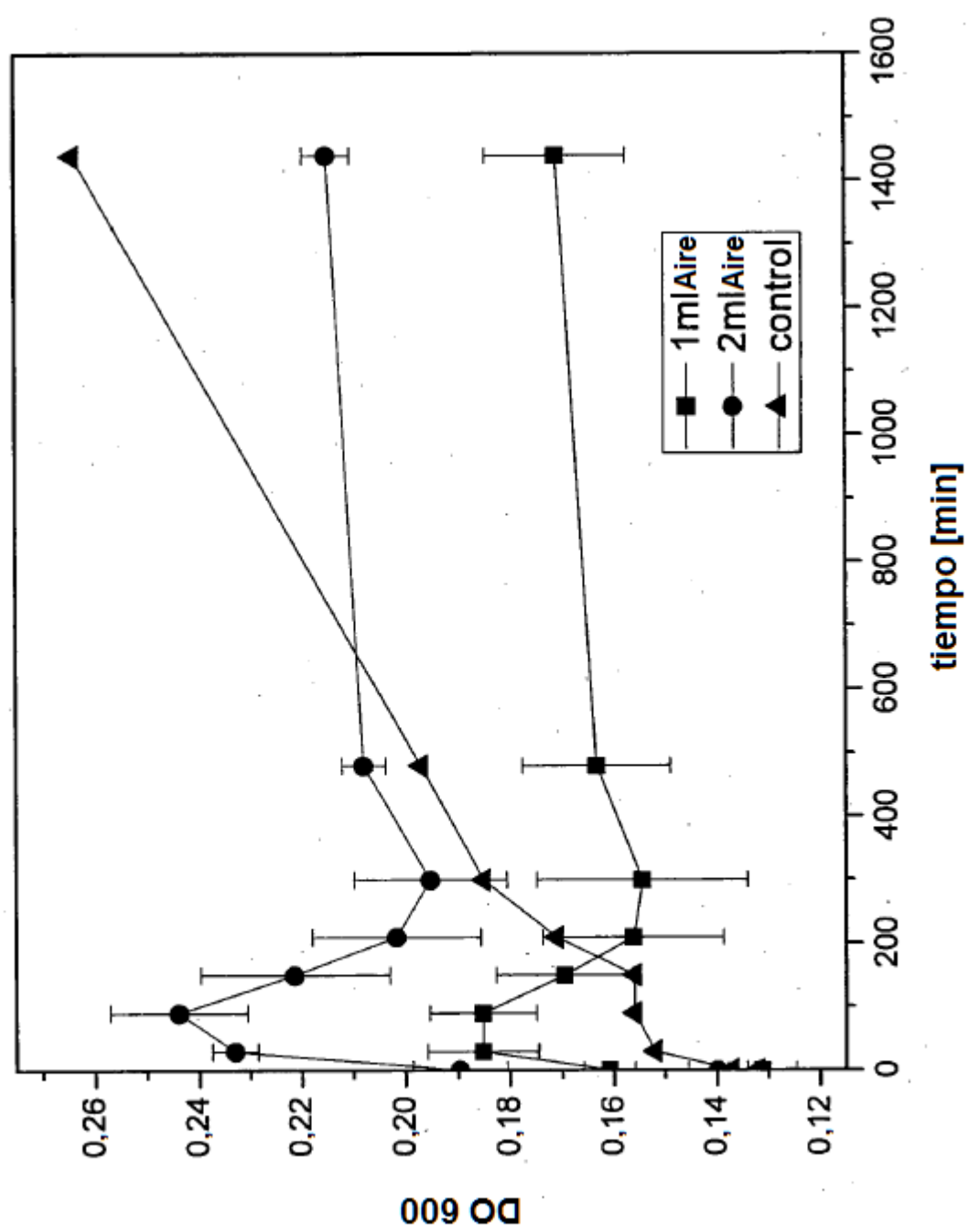
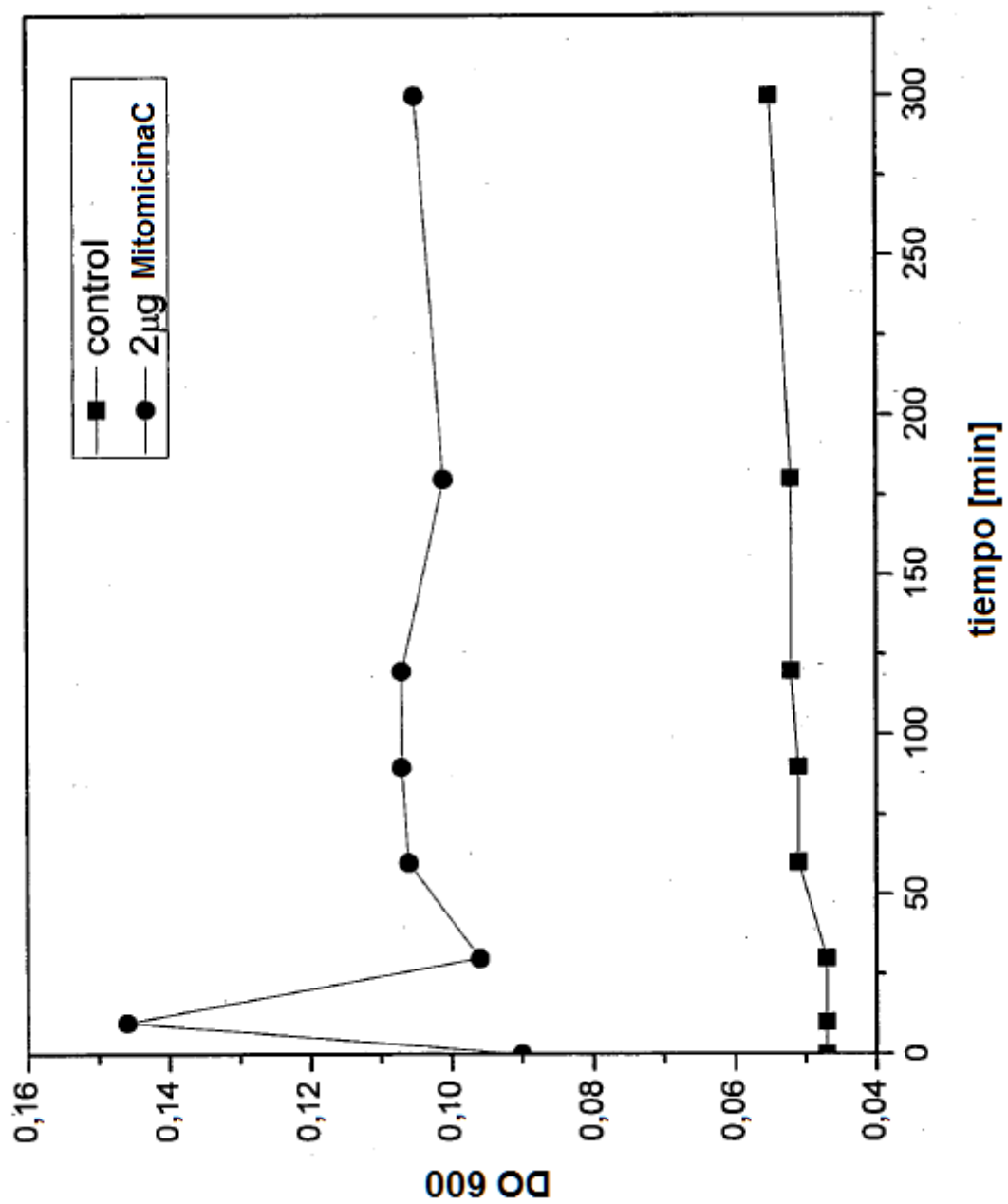


FIG. 1D



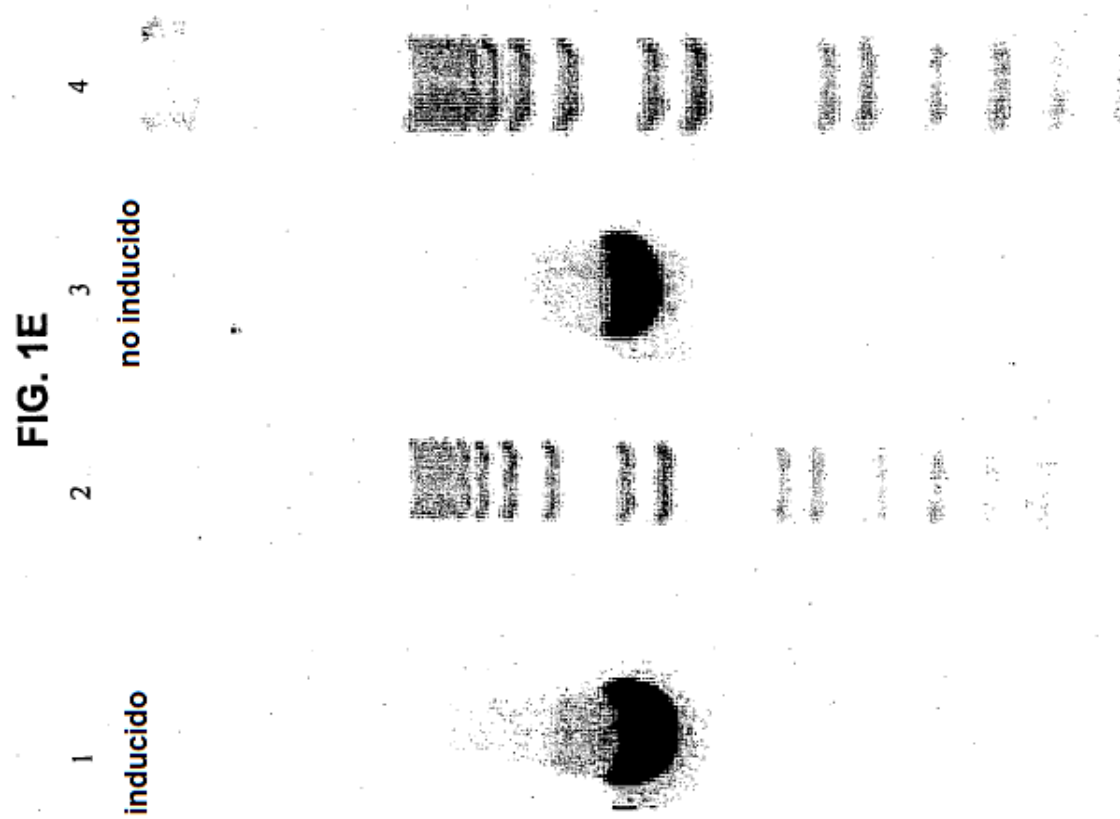


FIG. 2

Número ORF	Anotación	Comentario
896	posible mecanismo de infección abortiva, familia de proteasas CAAX amino terminal	mecanismo de defensa del fago
897+898	regulador transcripcional	probablemente actúa en ORF896, probablemente desplazamiento de marco
1451	integra específica de sitio relacionada con fago predicha	1450 muestra similitudes con una ADN primasa y 1452 revela similitudes con proteínas de fago y la sintenia del genoma sugiere agrupación génica, apariencia global de un elemento genético móvil
1572	posible mecanismo de infección abortiva, familia de proteasas CAAX amino terminal	mecanismo de defensa del fago
1578	posible mecanismo de infección abortiva, familia de proteasas CAAX amino terminal	mecanismo de defensa del fago
1579	posible mecanismo de infección abortiva, familia de proteasas CAAX amino terminal	mecanismo de defensa del fago
1783	proteína estructural de fago	muchos genes en la agrupación 1776 a 1789 revelan similitudes con proteínas del fago. 1989 está localizado en límite de cóntigo, que podría indicar una continuación de este fago remanente
1774+1775	similar a subunidad GTPasa de endonucleasa de restricción	ambos genes se asemejan a parte de un sistema R/M (específico de fago); la agrupación génica 1772 a 1775 predicha está implicada en la defensa del fago, sobre la base de sintenia génica
1959	posible mecanismo de infección abortiva, familia de proteasas CAAX amino terminal	mecanismo de defensa del fago

FIG. 3-1

ORF GAMOLA	Anotación	Función Predicha	Comentario
Región genómica flanqueante, locus de integración genómica			
1980	familia de prot. grande rica en asn/thr	adhesión celular	por integración de fago proteína de la superficie disrumpida. El extremo N corresponde a ORF 2069
	<i>attL</i>	núcleo en negrita	Tttttaattgattgtaataaattattctcgggtcgcac (SEQ ID NO:143)
Integración Cambio Lítico/Lisogénico			
1981	integrasa	integrasa de fago	
1983	hipo cons.		
1984	hipo cons.	interacción con ADN	probablemente alberga un dominio de unión de cinc y parte de cambio lítico-lisogénico
1985	proteína de control de la división celular	interacción con ADN	probablemente parte de cambio lítico-lisogénico
Replicación del ADN			
1988	hipo cons.	replicación de ADN	ADN helicasa o exonucleasa, similitud
1989	hipo cons.	replicación de ADN	
1990	desconocido	replicación de ADN	
1991	desconocido	replicación de ADN	
1992	desconocido	replicación de ADN	1 TMH (Hélice transmembrana) predicha, probablemente proteína de membrana
1993	exonucleasa	replicación de ADN	similitud a exonucleasas
1995	hipo cons.	replicación de ADN	
1996	hipo cons.	replicación de ADN	
1998	ADN helicasa dependiente de ATP	replicación de ADN	dominio de ATPasa implicado en el control de la replicación
1999	hipo cons.	replicación de ADN	

FIG. 3-2

ORF GAMOLA	Anotación	Función Predicha	Comentario
	estructura de terminador	se parece a origen de replicación	AAACCTTTATGGAAG TCGAGTAAGTATTGGTACGTATTATA AGTA TAAATACGTATCCTATACTTACTTTA TTAATAATTT (SEQ ID NO:144)
2000	hipo cons.	replicación de ADN	
2001	desconocido	replicación de ADN	
2002	hipo cons.	replicación de ADN	
2003	desconocido	replicación de ADN	
2004	desconocido	replicación de ADN	
2005	proteína asociada a fago	replicación de ADN	
2006	hipo cons.	dirección ANTISENTIDO	estructura de terminados adyacente a 2007
2007	hipo cons.	replicación de ADN	2 TMH predichas; probablemente sea una proteína de membrana citoplásmica
2008	permeasa	replicación de ADN	2 TMH predichas; similitud a la superfamilia permeasa; 2007 y 2008 potencialmente un desplazamiento de marco
2010	proteína de semejanza a ParB	replicación de ADN	escinde ADN monocaténario y mella ADN plasmídico
2012	hipo cons.	replicación de ADN	superenrollado en regiones ricas en AT
2013	proteína integral de membrana	replicación de ADN	5 TMH predichas;
Encapsidación del ADN			
2014	hipo cons.	encapsidación de ADN	
2016	terminasa, subunidad grande	encapsidación de ADN	subunidad pequeña de terminasa, en competición con 2015
2017	hipo cons.	dirección ANTISENTIDO	
2018	proteína portal	encapsidación de ADN	
Módulo de cápside del fago			
2019	proteína menor de cabeza	módulo estructural	proteína de morfogénesis de cabeza
2021	hipo cons.	módulo estructural	
2022	proteína asociada a fago	módulo estructural	también similitudes con ATPasa

FIG. 3-3

ORF GAMOLA	Anotación	Función Predicha	Comentario
2023	desconocido	módulo estructural	
2024	hipo cons.	módulo estructural	
2025	hipo cons.	módulo estructural	1 TMH predicha
2026	desconocido	módulo estructural	
2028	hipo cons.	módulo estructural	
2029	desconocido	módulo estructural	
2031	proteína de soporte	módulo estructural	similitud con dominio de proteína de soporte
2032	hipo cons.	módulo estructural	
2033	desconocido	módulo estructural	
2034	desconocido	módulo estructural	
2035	hipo cons.	módulo estructural	
2036	desconocido	módulo estructural	
2037	desconocido	módulo estructural	
2038	adaptador cabeza-cola de fago	dirección ANTISENTIDO	similitud con dominio adaptador
2039	proteína relacionada con fago	dirección ANTISENTIDO	1 TMH predicha
2040	proteína de cápside de fago	módulo estructural	similitud con proteína estructural
2041	proteína de cápside de fago mayor	módulo estructural	
2042	hipo cons.	módulo estructural	1 TMH predicha, secuencia señal potencial
2043	hipo cons.	módulo estructural	
2044	proteína relacionada con fago	módulo estructural	
2045	hipo cons.	módulo estructural	
2046	proteína de superficie de unión a colágeno Cna_B	módulo estructural	predicha en acoplamiento de fago
2047	hipo cons.	módulo estructural	
2049	proteína de medida de cola de fago	módulo estructural	similitudes, desplazamiento de marco potencial con ORF2051
2052	proteína de cola de fago	módulo estructural	componente de la cola de fago

FIG. 3-4

ORF GAMOLA	Anotación	Función Predicha	Comentario
2053	proteína estructural de fago menor	módulo estructural	
Lisis			
2055	proteína de superficie relacionada con fago	Lisis	similitudes con intiminas/papel potencial como holina basada en sintenia
2058	peptidasa		predicción como lisina sobre la base de localización/sintenia
Conversión lisogénica			
2061	proteína 3 asociada con el sistema dnd	sistema de modificación de ADN con azufre, dnd (degradación durante electroforesis)	
2062	proteína 1 asociada con el sistema dnd	sistema de modificación de ADN con azufre, dnd (degradación durante electroforesis)	
2063	proteína 2 asociada con el sistema dnd	sistema de modificación de ADN con azufre, dnd (degradación durante electroforesis)	
2064	ADN metiltransferasa de adenina restricción tipo II	metiltransferasa	
Sitio de integración del fago, locus de integración genómica			
	<i>attR</i>		Cagaataaggataaataatttaattatattgttttttttt (SEQ ID NO: 145), parte de estructura de terminador
2069	familia de prot. grande rica en asn/thr	adhesión celular	por integración de fago proteína de superficie disrumpida. El extremo C corresponde a ORF 1980

FIG. 4A-1

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
Región genómica flanqueante, locus de integración genómica		
integración cambio lítico/lisogénico		
1980	gtgagaaatatgaagaataagaggtttaataattctttattactgattacaataaataagcataggatctgttgtgcaacgga taatgaagaaataataggataataataataattgataataatgaggatatacgtaataattgataaattgctgataaact ccaatataaacaatccaactgacataagaataagacaattcaaacctaaatagagaaacagaactagattcaaaatttaataaatact	74
1981	atgaagacttttagacaacagctatttgaagagacctgaattccagaactacctgttgcagagagaccgaacctgacggagagcagccctgca gtcataacctcaatgccgccaccacacttcgtgaggttcacagggagccgttctacaagaccgtgcatgagctcagaagccagcagaacg ataggattgagaacaacatcatcataaggttcaaccggaaccagtcgaagataaaacatcatgcagtttgagttcata	75
1983	atgaatacatataaaatagaatgttcaaaaagatgattatcataaaagactgcaaaggccaaattctgaaaaatgagttgaaaagttgaagt tcctgagaactggaactgcgactatgtgaatgctgctgtggaagaggacatctgcgaagtgccttgaagagggtgatgaaaagaatgc ttttgattcctatgtgcggtgaactgcttctgaaggagtcgaggagatgaatacacaataaagtcacatttgcccttct	76
1984	atgggaatgaaaaagttaaagatcgaaaagtaagcaaaaaataatgacaagaaacgaaaaactcttctataaattatatgacgattt atataatgaagatagattaaattttattctccacataatttttaacatatatgataaatttttttaaaaaagatatataatccattatgtt taatgaactactctgaaaatgaaattattgaagcaatgaaaaaattgatgaaattcagtcagaaggcatttgatatt	77
1985	atgaaaaacacaagatcctaatttaataataataatgatagggaatctcctgtattcctcaacaggggaagcttttgagatggactatgtgcc ggacatctacaaatacagggacgagcagctggcgaaatggcgatgtactgcaattcaatacctgacaacatagctcccaagaacctgc aattgtcggaggcaatcgacagggaaagaccacaacataaaagcagttcttcaagatgttgaaacgaggcttttcca	78
Replicación de ADN		
1988	atgttttttgaaaaatgcgatggaagaagatacaatcagcatgtctcttcaagaaaaagatgaatctttatatattggaaactatggaaaagcaa aggaagcccttcataatcctgtattcattgtaaacataacctataaaatgaagcagaagaatggtacaaaaatggagaaataggagatcagg actttataaatttctatgatgatgtaaaacttgattgaagaaaagttttgggttgaaatttataaaaaaatcagaatat	79
1989	ttgataaatttcccttaaaactaaaaatttggacgtgtaatcatggaaaaatataatcgaaaattgacggagttcaatacacagaagcccaaat cagaagggcttttagccattgaacgagacgtgcgctctcctaatttcgttgatgtattgttggaagatcaaacctagcgaacttgcaa gcagagtttctgaaaaaggtgatgcctaa	80
1990	atgtgctatgtgggaaacaccaggacattggctctatcacacagaggactgcttctgcaaccactggctgttgaaacgagaacaagaccat tctagaggaaaagcctgtagacatgaaaccttgacgcttctgcaaacacacagtttgacactgaatag	81
1991	gtgttgatattggttcagagatggttgaaaacatcaggaaaaggagaagagatggatatctccatctaccctcttttccctgcattgt tttccacaaaggcaaaaagtattccgaatgctgctgcaaggctgaggaaggatcagaagtttgcaactgtgaacctgttaggcttca gaagagaggtatgtcaaggtcatagaccacaaggaccacaggaggtgttcgtatga	82

FIG. 4A-2

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
1992	atgagcatgcttgagacttcgagcctgcaaggctccacaagagacatgggctgaaaggcatgatgacgaatccttgctgcatatg ccttgccataagcattgcaatgctactcttcttgcgcttcgagagcagactgttgagagtgatttag	83
1993	atgactaaggaaattgaggacttcatagagaaggaacacaggactgctgtttcttgagatgggacacagtgagcatatctgaaatacct tgagtcacattatgacgaaaggaaatcgactacagatgttgaggcaatcgacaatctcttgacttctaccagataacct tcagcaaaaagactgaaaggaaattcctgaccagcttttgaaaaggacaagataaacaaggcttctcttcacac	84
1995	atgaatgagatagttactacaactaacgagaataatgtacctgtcgacgttgactatgccatcgaggaatggaaggcataccagcgatt gaccagagagctattggacgaaacagattatcacacacagagcaggaagtacaagaccaagagcgttgacgaagaatcgacacgtg ccttcaacataaacacacagataatcgacaaggaaatcgtcaagaatgacaagggaatcgtcattgaggctgagtat	85
1996	atgattgagtgcatcacaggaaggagacttcacggattggagggttccttctcctcttcggacatcaagtacatagtttcagtgga tgatgagggaacacctgtctgcagctgtcctgacttctattacaggaagtcaggatgaacctcacatctcaaacctgaatcctatt gcaaacacatcagacaagctcttgaggaggacaatagattgcaaatgctctag	86
1998	atggataataataaagacaaaaaagtttagctaagtttgaagagtttttcagcactgtatacaaggacgaagtcacggaagtttt ggaaaagtatcctgaggaaggacattggttggaactatgagaatttgaaaatgttgactgttgagcagactgttgattgaaa agcctgacgaagtatcgagcttcacagaaagcaatcaagaacattgacctattgatgaagatccaaagttggac	87
1999	atggctaacgaagttagagttaatctcacggttgacctaatgttggaattagcaaaagacaagttaccttgacgagaagcggaatt cttgagaatcagttgaagatgttcttaggttagagatgacgagtcggaatcattaagagacatccaaaccaaagaaatgagatta acgcactaagagacaagttatgtcatgttcgcaatcaaaagcagttgaaattggaatcaacaacatctatggagaag	88
2000	atgacaattggaagtttgacaattttggaagtcagatggagaaaacatgaacctgaagatttcgactgttctgtttctttgaaat gtcaaggcacttttgagataattgagttggaagttggcagcttcgctgaactcttggtgtgtacaagaatgtcgagatggattaca ccttgaaagagacatgccctcaagcagagaagaaattcttactggttcaatccgactggaagggaactttggaaaag	89
2001	atgaatgtgaaaacagtcataatgacctatagattgtcaaggagttcgaggagtggaattgagattgaatccaaaaattcaat ctacttctattcctttccaaagtacatgaaggaaaggcattgtgattctgaaatattccgcaatctatgacctgcacacatattgaaag gcatggtggaatcatagtgacattctggaggttgaggacaatcctggtgatgagaagaaggtttgctatatgtg	90
2002	atgattagcgacgaatgggaagaggaatactatgtcaaggtcaacgaggaactgagcaggtggaagtcaggttttcagcaaggctga caggctcgtcttgcataactcctactcttcttcttcttgcgagaggtgaaatcctgtgcacaggtatgcgacatac tgacaaacgatttgggaatgccaaagtattatctaggggaaaagatgaattatag	91
2003	atgaattatagggaatgggttgcaagccagtttcgagctgaggtgaggtgactgagggcgaagggtttgcttcaagacaccatgcagctacta cacattttcaaggaggactttgagataataaggagatgttcttcaatttcgagtcctga	92
2004	atggaagaataaaggaaattgaaatgtatgcggaacctgcaagcacagccatttgaattccagacatcaacggagagatagcagtgaaacat ttgcaggataggttcagaggtgtcaacaaggacggaggttacataattgcgtagactggacaccaaagggaaggtgtcatga	93
2005	atgttaagtaaaaaaagccatacagatgacgctggacaatgagaagcactatcctgtt aagtgcaagtactcggaagacggttccaaagtctcaacaacggcagatgtactgttca gacagttgcagacggaatgccttgagggaacagaaggcaagataaccaggctaaaaaggagg	94

FIG. 4A-3

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
	ctaaaaataaagcagaaaagtgctgattgtagatgaatacaagaataacggtttgggtagc tatggaacaaagtcgaacgcgacacagaaagaacaatttttccatgagtagacatggctatt caaaaggagatgaaaagaataggattgaagagataa	
2006	atgtatttggcctaaattttgtcctaactgtggaataaagtagaagaaaatgataaaattctgcatttattgtggaataaaactaaagagt tataattcctgaaaaaaagtgaagaaagctcaaatagtagtaataatgatgaaaaaactagcaaatatgttgaagtcattgatgggctaa tgagataaaagttttccatcaagtttacctgtgaataataattattataaagtgaatttatggaacaacatctgatgaaata	95
2007	atgatttatgataaaagcgacagttacaagcctcatagtggctatcctattggcatttgatgagcatgttaggaatcggagaatttgactca gaactacataattggcaatagtgagcgttatgatagccttgggtttggtactacaacgagaagcacaacagcgatttggtagcggaa ccaccaagtgcgaactgcgaacttgcctatggggagacgatgaagcacttatatga	96
2008	atgaagcacttatatgagataaatcccttacaggaggactgtttggattacaggggttccctgaagacaaacagtcctcctcggaatgataaac cactggcgtagtgtatattgttcaacagcataaccgagcacccttacttcatggagtgggacgagataggaatcgttcttggaaatcgttt ccatcaccatcgccctgcatttacatagcgtatagacagatggaaggaacgcaggaagaaggaggttgacacg	97
2010	atgatagaataatcagcaccataaaagatcaccgacataaaagcctgcgaatacaaatccgaggataatgagccagcttgaacacacaaagct caggaactccatggagacattcggagtgttgacccaataataataacacctgaagaacaacacatcataggaaggcaccacaaaggtagc aggctcttctggacaagtccatggaggacaacgagttcataaaggagctccactcatcaggcttggagatgtaggc	98
2012	atggacgtcgacaacctgctgggatatgacaaggtactggaattgaggtcttatttggaggaggtcacggacaaggtgatacccgtagt gcacaagaacccggaaatcaaggacttcaagcagatgtgccaggactacaatttctgcagcataaagcggctggagaaacgaggacgtca aggatgaccagttccattctcgtcaggcatgcacacaggaacgctgcaggattcacggattgggattgacacggagaaagggtctc gacagggttcccttcgacagcgtcgacagcagcagctggctccagaccataattgtatgcaagg	99
2013	atgattaatgaaggataaagaccttatttgaatttttagcttcaatgacaagaagtcatttctgcacattgttctgtgagcaatct gataagcaacctgttagccatcaaggtttcaacttgggatttggggattgacaacccgatttgcggaataatcctctgttcccgtaggat accttatggcagacgtgattacagaaagtctatggagagaggacagctcgaaagggtcataattgcttgggctcttttga	100
Encapsidación del ADN		
2014	atgcccgaaaccttgggaaggcgaaggacgagaacggaaagcttgagccaataaaaggcatctcgagtaacttccaccgagtagcctgacaat ggacaagccacgaagcatgaggttctgtgcgagagctcggcaagaagaatgggtatattaggcaacttcatgcctactcatccacat ggaactgggtcgaaaggcgagagccatagcagacacataatttgaagaacgggttaaggaaaagagagattttac	101
2016	gtgaaggacattgtcaatcattacggctacctttccacctacaagccgacatcaggtctgatagtaagggttaaaacaagttcaagct caatgagccttatcgtggacagatgttgagtcaggtcgtggcgttcaatcatggatcaggtgcaggtctccttatttgcgatgacc ctatcaagaacgctcgtgaagccgaatccaaggtacgtcaggcaaaagctcaaggactggtggggagggaaccatcaag	102
2017	atgagtaagaaacaagaatgatgagaattgaaaggttaaacactatgcataatcaaacaggactaataatccctatttttaaaaaatca tgatttaattaaaaaataagaaaatgaaaaataactactgacgaaataaagatatatatagggaataatgaaagcagataaaaa gaagaagagaaatttataagttatttattgataattgcaaatattttaaatacagacagcatatgttataaaataatataatccaaagtaaat	103

FIG. 4A-4

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
	aatttt	
2018	ttgagtaagaatgctaagcagacgcctttgtgtgaccactgaagacggatcctatgacattgtcgtatgcagatgtcttgagagagta tgcaataaagagcgaatctgacgagacaggaagcaagcagttgaagactgatggttggaatatgatgatacatgtcttgagccgttgt atgacccattgcagttgtgtgagctattggagataaacacatatcatgagaactgtgtcgtatgtgttgcgaaggac	104
	Módulo de cápside del fago	
2019	atgatggcgctcaaggtgcgtcattcaagagggcagattcagaatatcaagagagaatacagaaggcgttttgatactcgaagagcagtg cagtagggagattgcggacttcttcagaaggcttgagaggaaaatccacaaggtcatggatgagcactgggaagcgaattgggtcctt tccatctgaacaaggttccagacatcatacaggacagcaggaggtactatgacatcctgttcaagtactgc	105
2021	atgataattgaaattcctggaaatgacaggacagagaataaattgcccgaacggacagtttctactgataaacctactgcaggaaaa tgacatgatgtcattgcctgacgggaagtatgtctgccattcagaatgattcagcttttcaactgaagggcgtgaaactgattggag agtgcatggaggaaaatcctcactacaacactgcctctacaatactgtctttgagcatttggatgaaggaaattgag	106
2022	atgttaagagtcgtagagagaacctattatcaacaagaagaatacaagacactagactgtcgaattagagggcaggtcaacactta ctcccttgctcgtcaaggagcagttgactatcctacctacgaatcctacaacgagtgagagaggaataaaaggcctaaggaaa aatacgtgaaatcctactactatcactggagagacgttgagacctacttctattacttaggaagattctttagtgat	107
2023	atggacgcgataaacgttatcaacaaaacaagatatattagtagatgtttgtacctgggactgtaaacctgattgacatcgtcatcgg tgatgcattggtctatgacaatccgacagttgtagtcaagtgtcacaccacagatttggccttttgcaccacagagattgatgagattg tattagagaacgaagaatttgagttattataaacttacaggacggagagttctcaattttgttgaagtcaccacaat	108
2024	atggtttttgtcgtatttgcgagaaacgaataagcagacatggtcagagccttctatcgaatggagacggtagcagcagatattgac caacagagcagttggagagatagtgagcactactgttctcttgaagtagacggtagaatcaccactaatctccgagatttccctattat atagttgtcttatatgacaccattggagaggtgtcgtatgacggagtcacacttgaggaaagtgttcatcatagaa	109
2025	gtggtgaattggttgaggctattggagataaacttctcagtggtattatctccttttagcattattctctagtggagatttgatattggt tgccatcgttctcaattcttatggtgtcatctcacctgagaatgtaagggaactggtgattgactataataagttacaggaaaagtgtgata tttttggagacatctaaggagacctaagaatgagtttgaagattatgtcttagacaattttgaagagatggaaact	110
2026	atgacaatgaaagltacttttgaagatgagaacggtgagagaacagtcgagttcggcgatgacgtcgatttctgttttaatcgaagcga tgatgacggaaacattgagattcgcgaaggagactgggaattggacggagattccgacgatgattgggaagagtaacgatgactgggacg aggaggagttctga	111
	gtggacgaattcgttgagacccctgttcgacacctactggaaggtgaatgagaacggagagtagacatgagccttaccgattgcggtgattt ctacatcgcaaaaggttgctccctgtgttcgcaactggagcatagtagatagaatgcaattgcttctgttttctattgcaaggaatttgcct accatgagaacggagcaataacttgagataggtatggagattagctctcttacctaaagccaaatggagataaaaagatttgaagatatac atgatagattgt	112
2028	atgttttttggaaaggaatatttgaacaagacggagagagaatgtaagagaaacaggtgatttactggagaaaagtcacatcaggtacacaactg gttcgtagtcaacgctcaggacggagaggaacaactgccaaaccacactcagtaagcagagagaacagtttagaggaactaagagacctatgca	113

FIG. 4A-5

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
2029	gagcagtgctcgagacaatgacaaggcagaggaactctcccaacaagacacctgtttctctttgtgcgatagac	114
2031	ttgcagtttggtgtagaaggtgagaatatggaatgcccttgcgataattgtgagatgttaagagagaacgaaccaatcaagtgattga ctggaagtccaattctccatttggaaatgggatttktattcacttttggaggaaacgcttcaggatgttggctataatgggtga	115
2032	atggaactgatgactagggaatggaagggaactcaaatccttctcttcttcccaagacggaagagatgacgcgatatggt gatgaagtcttcaacccttacggattgggaacttggatgtattggaggcagagaaaacaggagaatggtgattacctttcttcggat atgttgaatcacggataaacacctgaattcaacgagtagcgttacttctcattgtctgagttggagaaccttaagata	116
2033	atgttcggacagaagaatgttgcataaagatgtatcatgtcggagacgttgcgaactggttcataatggacgatgcacaggctcc acctagtgaaccagagagaataactcttcgtagacgataattggtcaaatccatgtgagatgggagaacggtcaggctcgtgcactaa tatatggtgagacaggttcaaggtttagagaggaaggaatatag	117
2034	atggatttggtagatctttatgatgacctgactgcaaggagagaagtctacgattcagtcggattgagtttgggtcaagtataaatt ctcatctcgttaagggaagcacaggatttcgcattgaagtagcgttcgagaactgattgaggaatga	118
2035	atgattgaacggaggagattaggaatgaagtacgatatcttttactatcttggatgaaatttcaaggaaagttagatgacggtgaattatc cgatgagcaggttgatttctctgttcagatggagattcttagtagaggaaggcactatcacagacgagcaggcacaagatgtcatgaatg gagattattaa	119
2036	atgatgaaatgaagttgaaggaaactcgttgaggaaattgagatgatacttgcagaaggtggattaaacctttgatcaagtcgattatct gttgacttggagacctgcattgcagacggaagcataacagagagcagagaagaaattatctgtagggaacttctaa	120
2037	ttggtgagagaacagaggagattgattatgtatttggaaattgcagaggtcaagtcgaaacattgcagagaattgagtttgacgatgt cgcaatggaaatcgtcctgcagacagaaaaggtgtacgagagaataaaagagatggctcaagtccaatgagattgactacgattgcaggg aagaccggtatttccccaatctcatagaatatgtcattaggataacttgggtgtaa	121
2038	gtgtaaccgtgagagatgtcggattcaccattgaggaaacggttcttcttactgctcaggaattggagatttccggaagtgggagagga acacgagagcgtcattgacagagccatcgcatgtgtacacgaagctcagaaccagagacttcgagttcacccgacgaagaacgggaac tcttgaggatgcttttgcattgtcagcgaccaatag	122
2039	atggctaaagaaaatgttatagattataagattgagcgacaaaatgataaacacttggagttatttataatgtaactgaaagaggagagg aaatatcattgcttcatcttttgagaattgaggcagaaaagttttaaagagggttgccttggaatgatataatctaacctattcacta aaaagagttctgattctaattgttcgtcgcgaatctgtaagaatatattgatgatgaatcgttttggctgatgcgt	123
2040	atggtagtcaaatgtcctaattgttttagtcacgtgtttcaaaatgtgaagataactaataattaaatggcaatgtgataaatgcaaatg caaatttaaccatggtgcttctcgataattaatgctgaaatggagaagtagaacaattaaacaatagaaaaattgaaagagaacgagaaa gaactgaacaatttgaagagctattaaagaagcaaaaggaacaatttgaagagagaagaactgaacaatttgaaga	124
2041	ttgaccgaggtcagataaaggcagattgcacatgagtatcttgcaaatattacagccttgcgacaagaacctgagttcttcgagaccag agaagttatagaggttccgtgtgagtcataataaacaacagcctataagcctcaaggcttgacggaacagtcgaatgagtatccaa aaggaaacctggatagccaccacaaggataactgatgaagaggagatggaaaaggcactcaatggagagtagactgga atgggaaacgaagctactttaaccatgttgtaaacgagcaggaagggcagttcttcaagtcctatgaggaactgacatggagacaggaaa ggctgtattgaacgtcgagcagctaggttacttctttagagaagcaacatttagacaatacaaattttaagagatcggaacttcaagctga	125

FIG. 4A-6

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
	tgaatcattcaagaagcatctcaacaggttaggaataaaccggaagggtcctcacaacggatatgacgtgaacggcgagaccgcacctgagattcctgcggctgacgtgcacttcggagcaaacgagttggacgtcaagaagctcaaggca	
2042	atgcggaggcgttgccctgaactctccagagcataaattggtatgatttcgcataatcatttgctctctcatttgcctttatagagattggttga ggcgatactgattggcattggttattgattgggaggaacttggcaaatatccgttcgcaagctctcctagaagcctttataatggttttgaaggatg agttaggctcttcctgaatggaacgaactgctgtgattgaaaggagaagcatgaagaaaaggtatgctgtcgcataaga	126
2043	atgacttgattggtacggaagatgtcatcgagttcacaggagtcgaagcctcagacattcaggttcgagaaggagagacacctccagcct ggaacattgcttgagaagtggattctgcaggcagaaggacttataaatatcctactgcaactacgatttcaatgacttggaggagatac ctccagcagttgttaattgtctgccttaggctcactgcaaacatggtcgcattggcacaggcaaggagacactcct	127
2044	ttggtgaaattgcagattgacgttgaggaactcaagccattggagccaaagttcaaaaaggttgcacaaaggacagattgtgttgactgc aaatgaattgcagagaaaacctcaagaagttgagtcctgtggtatcatggaaggcttcagggtcctcttggttaattctccagacaggagaat tggaaaggactgtgaaaagcagtgcaaaagtatgccattttcgtgaatgacggaacaggactgtacggtcctttgggtcat	128
2045	atgagattcgttaataactgcttcgcttgcctcagactgtcaaggcatatcttgaaagggaatctcgaaaggaggttgcttgagga tgtagagacactcattccgtccgtgaacagcgacgttccgttgaccacctcgcatatggatagtcagcagccactaccagatggt ctggcagtcagccaaatctctcaacaagatagctatgtcagtcctcttcgagttcgtatcggtggaatacagcgatgacttggaaagag gctgagatattgggaataagcctagccagcagggttggtcgaagcctgatgaagaacttcaacaag	129
2046	atgggaattcgtgttaggaatgaaggaaaggaggaagggtatggagttcgaggtcagcgcggacttccatcaggaagttagcaaggc aaaggcttccttgaactccactccgaacacaaagtcaagcggtcgaaggatgaagaagaaggcacgtgcaggcgtgtacaagcctactg ccaacatcgaaagtgaaagtgacttgaaaggagataggacattatctcaagggttctcctggacaattaccattttact	130
2047	atggttgtagtaagaaatccgatatattttaagggtcgtaaaaaagattgaaaaggtgaagattgaggcttggacggagacgagatgta cttgagaccgtttgtcccaagccgagatcaacgaggttcgacgagattgaggcaaaaggctatgggaatcttcgagaccaacgagaccgcac acaggggaagaaaggcagaagcctaagagtggtggttgagagcaaaaggaaagataaatctcgaattgcagcagaaggca	131
2049	ttgccgtcaagcaatgtaatagaacataatagtcgaaggcagaggaacatggcatcatctgtcgcacaaagggttgaaaacagcttcaggaa attgggaaacacaatagacagcaccttccacacatctcttcaaacaccaaagttcaatcaggaactcacttcttttggaaacagacttgg acaaggtaacccaaaggctgaagcaggtgagcgtgaaacggtcagctccagcgttcaacagcttacaataatgctgaaaagg	132
2052	atggcaagtgttacaaaaatataccgagcaacggttccagagaccactggaggaaggttcgtgtccttcagcaaatctggcaaacataaagaa caatgctgacggagcgcattgccgtgagcagtggttcttatcaaaaagcaagaagcagctctccaaacagaccgtccacagctctcatgcaaa gcttcggattcagccttccgtgaaggtgctgaacctactaaaaatcacagtaacttataggcataggaaagatgctggaagcgactacagc tcaagaacaagactcacatatgcaacatcgaggagacctaact	133
2053	atgggcatagctattgttggatggacaatgaggagaacttccctgcaattccttgaccctgatttatgcaccatcaatgagaccataga ggaattgggcttgaggactttggagttcaactacaagtttcaggactatgttgaggacagggatcttttccaggataggtaacaagatctt ggattagcaattcccagagcttggaggactgcctttatgtgataaacactcctgtggagaacagcgtctcatcaggag	134
Lisis		

FIG. 4A-7

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
2055	atggtcgagaagattacagtgagtcctcagggaagtcagaggatcacggaatgtgtgtgatgagaaggaattggaagattacggaagcta caggtgtgatgtgagcgagagctcggaggtgatcaaggaggttgaggaaggaatttcagcgtttctggtgttcctgctcgtctga gcattgccaatgttactgaggacactcgcaggcagggtgcgccacatctccgctcatttgaggatggggaaggc	135
2058	atggttagattcagcagagacatgctccaggcaggcgaagagaatgttcaagtggcgaaggaaggcgaagggttgcctaactactt gataatgtatgacatggacaggaataaggagataaagtgggtccaaaggaaatgacaggactgtatgagtcagaaacataattctgga ttaagaaagggaaggagcctaactatgttacactgactccgttgcaaggaaatcctctgtgatggactaccagaaacacaaattacacc tggtggcccaaccagtttgccttgccctcacaaatgctatatcactataagtcagaaagtgaatgcgctaaggcctttgggaactagcaa ggcagtggaacaaggccctgcacag	136
Conversión lisogénica		
2061	atgaaaaaacattgcttttttttaggggacagttttgccgatatatgcaatgaagcc atgtttttagtaaaacatttagtagaaggaacatttttgattcaattatccgtgcagga aaggcttcagagataataactgtaaacattttggaacttgagggtcaaggtgaccttaata agctctggtcagaagaaagattggaaatgcttgatataagggtatcatttcttatgat atttataaagattaaaccatattcgtaaaattagaaataaggctgtccatgggcatttaagtgtatattgaagacaatgcaaatattct gcatgctatctgtatctaatgtgcgcatatttctataaggaaatagagacactaattttcagcggaagattatacgggcccattta tgacattgacctaaagcctaagagagactgcttcagagacttcagaggataatgaaataattggagagttatttcaagtcattggat gattatcttttgaaaagatgatgacagttacctgttaaatgaactgtctaaacttaaggactcttcaaggaaagctgttgaagacga taacttaagcgaatttaaggaaatcttcatattgacagatctattcaagaaagattttttaaaggcattgaacagagccactagtttta attcatctcatctgattatgctttgcggtagttggagatggaaatcacacttgatgcaaatttaaaaagaaaaaacctgaactc tttaaatcaatttgctatccattatgatgctacagagagttttgatcctgaaaagaatgcaatagacactttagcctcagatttggaaacc tttcaatgacaataatcttaacaattcgacagaaaaacttatttagccattaaacttggtgtattgaacaatttccctggaatcatctt atgctaataagagattacactaaactcaagtttaattatagaagaagcgaacataatttgaatctaattgaagtttcagataaacatttatgga gataagggttagtttgcactttcagtgattacaacatgtttgaattgaatgatgataaaattcaaacacacacatcaaaagtacat ctctccctattttaataagataacccaaaggaggatacaaaccttttttatgttgcatatctcaaggataaggactctcactttatca atcctataatctataatgatgatgctgatgaggtccagaagacaatcattgattatttaatacagattttcataaaaatac agaaaaatcatttcaactagagacttattgaacttcatattatgagataatagttcccggaattcctgaaaagtggagatttggacaa tatcaatgatttcagactattcattgcctaatttactcttcggatatccagaaggtcagatttattgaagctatgcaatgaattgg atccgacattgcatcgtaatgaatctttagataagttcataatagacttgaacattaatgacgatactgaaaagataattaaatcgttat tttgatttcaactcgttttaattcccttgaaagagatggagagatctggttgatttttagggagtttaacaattctgaaaaagagaaggt caccataattctaatagggttgagattctctatgaaagagcattatcaagaataatttcaaggataagggttatttaaatatttaa aatatctgtatgcctataataacacagctcataaggattacaataatctgttcactgaagttaaggacgctatttttaattggaaaggt tcctataagaagaataactatatgtattgacacttttgattcatttaagtgataaaaaatttaaaataaaaccatctgttgataaatt	137

FIG. 4A-8

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
	tgaacaaagtttattagatggtctcttttttaggaataagattttaaaccagatatataaaatattttttcagttgaatctaaacaaaga aaataccttttaaatgttgatttttcatatatataatcaatacaataatgaaattataatggttttaagcctaataatcaatcagataaagcagat ttaataatttttagatgaatttataataatttattagatgaagatacagatgatgatttatattgtaattagtttagaaaacttatgaaga attttatttgagtcacaaatgattttggcacttttgaatttaagaggggttaa	
2062	atggatttttcagaaaaattataataattattataaaacaaatgactttgtgatgataataaa aggaaatttaatacatcagatttaatacaaaattctccattgttaccttttaaaactaataact cctaaaaagctaaattttgaaaatggttttgatataataattgggtgaattatccagaatc ttattaaataaaaactattgaaaaaaattttaaattagataataattgtttcaaatttaatt gataataattgaaattgaagatggaactaaggagtgatatacaaaaaattgctaaatgaa tatttttcgatgaaaaaaatgatttaaaaatctctcatccgaatttgtatttatataatt cctcttcaataataaaaaagttcctaaggagacaggaagtagcattgtttttaagagat attttttgaagaataatcaaaatcttataaacttttttgaaatcataatgattcaaatcac attatcttaaatttaattttaaaaaatactcctaattttacatcataaaaataactgaaact aagtaatgaattcattttgaggaatttgcaaaatttatttaataagaatatcaattatgca atactttataagaaattttttatggaaaaacattgtgtaattttttgcttattactatttc ttttatatctcctaatttaattctaaaaatttctaaaggctttaatgataataatgaattgaaaaattgtactattttggactggga atctgcaagttaaaaaatcgtaaaatcattaaatagttatagtttactaaaaacatcatctaaaccattatatgcaaaaatggcagtcataag atcaataaaatacacatttaggtacaaataaatttattggaaaaagataattcagaataattttaataatttgacatcaattctaaaaat aatttttacattttcttaaaaaaatgggtttcagattataggtatgtaaggaaattttgatgataaagaattaccagataattttattaga attaaactgaaattctttttgaaagtttaaaaaatgaaaaactgggttggtgagctgacaaatctaggtacgcttttaaatctttgaag atatagctaaaaaaataattgttaagcgtagaggttcgtatggttatgtattaaataatataagagatattgttgccttttaactgca ttatgtgttaagataagaaaaattaaagttaaatcaactatttattgaatttgaaaaagagggagtttattttgacaaaattatcaaaaaga agaagtagtaaaattttttaactaaatttaaaatttaattgacaagaaaaagtgatgtaggagatgctcaaatatgttaaacccagtttttatgat	138
2063	atgttaaacccagttttatgattatctatctaataaaattactaaattatttttgatgataactaagatttttaagtggtgaaaaatttttcat tagttttgagaagaagcaaaaataatgtctttttataatagtttaagaagatttgcaaaaactaaattttctgttctgaattcattat atgtccatactatttcaggaaaggaataataaacttattccataaaacatttaagagtcacaaatttgctcatttctgaaagtttaacaatt aacgtagatttcttagttactttaagaaaatcaagttacttctcaagaggggtgtttggaaggatactgcattatttagttatctgtaaatga agcaatagatagatttgaaaaaggtatgagaaatttacagaaaaggaaggaatgccattaaatgtaaaactctatttcaaaaaatttagagg atgaaaataacgattctcaaatctttaactattccgatgaagcaaaatcgtaagttttcttttgaaatttcaagaagaagaattgtttcaa acaacattatgggattatgaaactatcttcaataatcaacaagggtttttgttagtgatgaagatttaagagagttgaatcttttttaa agatgaccaactaaatacaaaattctcctcaaaagatgttaaaaagattaaaaaataatgacacatttaataagataaaataaaattt cacaatatggagataaaaaaagaaacagcttaaaaatattgttactgattctggagtttctatttttaagtaaaagatgactgggtataaagca gaatggaaaaatggttaaaaaatcaaggatgactttatttaatacaaaaaatcctttaaatcaaatgaaaaatccttgaaaaaatcacaga	139

FIG. 4A-9

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
	aaatggcttaaatattgggaaattcctaacttactaaactggaaaacgaaacgaaataatcattgttttaataccaatcatt ctcatgaagttagtttaaaagtctagttttgatcagattcttagtaattcttcttaatacacaattcaaaaaattcactattgcaaga ggcaaatcattaataatatttactctggttagtgaaacaaattttcaaaactataaaacataaaaaatgaaaatattctc tgaatttacttttaataatattgttttgaaatttttgagcctgaaatttttaattctattaaatctcgttttagttcaaatctca aacagattattgttaacaaatgatgaggatagtttttgataattgtttttgactggttcaaaagaaatagagaaatttaaaagaaaat ggt'gaaaagtttatattcttatgatgatgaattcttaaatctgaacaatctcctgcttggaatgatgaaaaatttaagttttaaat atataaagacaataattatgccccatttttgattaaagaaaaatctaagaaaacacgtccagtaaatctctatgtaatgttgaaattgga aaagaagaaaatggaacacattttatttttaagtgaagaaagctgttcaagatgtaaacagtttctatctcgtagaagaatttaagaa tttctaaaaatggaacgagaaatcattaaacagacatatattttatgctaaaaggaattattgatggttcttttagaaaaaattgaagtttc attcagtaacgagttagaacactgcttacatggacatatatttaattattataagacttttgatgattctccagaagacaatcttccaaagt tagtttatttaaatgatgatttaaaagagttatataagaaatttataactatcttcaataaggaataatcagaaattgaagagaattct attttgctgatttttaaatataaaaagaatttgccttaagttaggtcgtattgaaactgataacaaaataatgtattctccattatctcc actaaatctcgttatcaattagaagtttctaaacaatcggaatgaggatttattccgttaaatatattagaacggttagttccaaata acttaattccttataattgttcagatgaggaagaattgttttagacctatttatacaagaagaagctcaatgaatggttgatttatgag aaaagtgaagaagtataatgttggaacaaacaaatgtatttatttctaattgttactgaaaaacctaataatcaattgtcacaacattttaa ttacttatttagttttaacaactcctcacaatttaaaattaaacttaataataaaagatgataaagaattgtgaaaggagatttta atthatttagatctagattgctgataaaaactaaaaaaaagttattcctgttgaaatcaatattttacaacgctgaaaaagagt tcttttgacaatttattgtgattgctgagttcgaattcagttactggaagaatttggaaattaaaaaattaaaaatcagatatattttgatcc tatagatattatccatattgattcaaaaataatttcatattataaaacatccatttataaaagagattatcttttaaacggtttattatcttca tctataaagtcaaatcacaatacaaatattgcaaatgacaatatggataaaaattgaaactggtatttataaacggtttattatcttca gttacatcaactactaaacattcagaatatagaacggttttcggtacaaaacataattttaaataatgagtaaatcctcttataaaaacagt gattaacttgaacgaacttgttgagaatagcaagaacttggaaaaataacttattcaaaaaataaatcagttatcactactgttgaat tagaagaagacaatatattgaagaattgtacgacaaatctcattgggttaacatttattgaacctacatttggatcgaataacttggacagt tctgatagtaatttgatcattattcattacagcgatcaatatagttcctctagcaagtatgatactattactgtactaataaatacaac acaatatgaggaaattataagagatttctcctccaattcaaatatgtaaaagttacagatgagaattgtatgatgtaataaagattgttta attcaattaatggtgaattggttacttactagagataatttctaatccggccattatagtagggaataaatttaagattatttctgctattaaa tattgcttattccattttggaccataaagataattgtttggattccagtttctatggaagaatacttaagaattgcaggaattgttaaat agataaaaaataaagggaatttttgactctaagtttaataaaaggaaatcatagtgatctgctattcatttggtaaaaaatccaatgaag ataatagaatcgaagttatattctacccaatcgaagttaaaattagtttgaataatgcttcaactataaaaaaggttaaaagtcaatta gataatacttataaaacttcttaaaactcaactcaaaaattataatgtagagaattctgaatttagaaaaaattcttttagaaaatttctt catccaaaattatttgcctcaatgagcaaaaatttagtaactaatcatatttgggatgaaaaggttggatagaaattgaagaattcaaaag cagaactattaaatgatgaatatgatatttttatattgttttagaagagtcatttggtaaggttcttttggtttcatttataaaaagaatct catcacatctctatctatggttggatgaagcaagtaattgagttaccagaagattttgcttattacggttttagccactcctataag agaaatcccatgatgagattcagtcagacaatacagatatattcttgcagaacattgttatccatgtagatataagtgaaattagagcaa	

FIG. 4A-10

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
	agaataatgatataatcgattcaaatgagatatgtcaatagatgatgttttgataaatttaagtgatttgaggatagttttattgaa gaagaatcagaaatttctgaagagcctgatgaagaatttaactgatgcaacttcaagcagtgataatgaatcaattgagaatattggaga atctccttctaaatttctaatgtcagagcatttaattggaactcagaagggtataatcataaagtttattgggaatttggacatcctt ctttagcaaataggcataatgttaattcaaggaaaaatctgacaaggtaaaacataatttcattcaacgaatgttaaaagagttgtctatt caaggcattcctagcataatattgattatactgattgattcaagccttctcaattagaacctaattttaagactctttaggagataa aataagccaataatttgttgaagaagaaatttcccaataaaatccatttaaacgaacctatcatgattgaaggacataatttattg aagaagacaacagcacatattgcgagccgattttaaattctatcatttaattctgtttatggctaggtattcaacagtcacatacttataat caaacagtttttagattgcttggataaatttaaatgattgataaatttggatttgaataatcctaaagaagaaaattctaaagacgaaatcaaatag tgcacaaaacagctcttgaataaaattaaatgagttattggataaaaaatccgttcgcactcactggtctgtcttagataaata agacggtaaaagtttacattatccaattaaactgcactttctaaaggataacagacaattataaactgaattcctttgggatttatgg aattacaaaattaaacaaatggtagtgaagataatcctttcattgtgttttagatgaagctcataatcttgatttttccaatgactctoc atgtagtaaaattcttaagaagcaggaaatttggatggtctgggtgtttgcaacacaaatctgttaaaggatctatgaaaaattgatg aaatagctaaattagagaatgcagatgagaagataatatttccatccaacggatgtttcaacaatagctaaagatttgcataaagacaat gaagataaaaaaatatataaaaaagagttatctcaattgacaaaaggatattgtattgttcaagggtctgcaatagattcaagtggcaa tttgtatcagccaaatcctgttaactgtaaaattggagagataaagttttgatgaaaattaa	140
2064	atggaagtaaaagacagaatagagaataactattttcaatgataaaagaatggatgcttat atagacaaaatttttgaagagtttactctgagttctgaacagaagagcactgaatatt tggatagataaaattaaataacgatcaattaaacagtgaaaagggaattatcataatttt tttgaaaattattcttgaagatttacttgggtataaaacgttctgtgtttaaactgaagag aatattggtgatgaaggccatcctgtgaggtttgtattagaaaaagatggaagaattat gtgattatagaactaaaaggacaactataaagacactcaactaaaagaagacctggacag caatcaccagtagagcaagctacaaattatgctagtgtctaaaaagaactgaatggct acagtttcaaatattgatgaatttagatttttttaattccaacagcaagagataattatatt tcattcaagtttaggcaattgaaagatttggaaatcttttaaaaaatttttagtattc agtaaattttcaacttatcgatgaagacatacctaaaaaattacttaataaaaaaaagtc attgaaagagaattagaaaaatgaattttatcaattgtatgtgacactagattaatgatt attaaaggaaattggaatatcttcagaagatatataataggattgaagcaataaaaaattatct caaataatattaaaatagatttatttcttctgttttgcagaagatttagcgcttatggag gaagagacaactgctgatgtattataacaccttttaaaagcatagaaatttaatacggaaat acaatgtggaataggttaaatgaattatttttttgcataatcaaggaaataaacatagg agaataacctgcatttaattgaggattattttaggaggtatttatctaatgtgaaaaataagg gatgaaattgaagatagggtcttcttcttgaattgaaattttaaagaagatttttgaagac aaatatgaggatatttgctaaactaatgtgaggtttataaagatactcttaataatccatttc	140

FIG. 4A-11

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
	attaatttattaataatttctacatatgactttgattcagaacttgatgtaaatatcttta ggacataatttttgaataacagattattagcgataattgaagagtttaaaaaatgataaatcaagag caaagaaaaaagatggagtgattacactccagaatatattacagattatattttagaga aatacaattattccatattttaagtatttctgtaagccagcacagttcatgaattatta tacgaatatgaatcatctaatcattggatgttttagattcaaaaataactaacatcaaa gttctagatcctgcctgtggtgtagtgggagcatgttaataaatcagtagatatattttt gaaattcatgaagcattacatgcaagtaaatatgctggagattcatcttagatagattt tttgtagattttagaaaaacgaaaggaaattataagtaataatttattg999ttgatttg aacgaggagtcctgtgaaattactaaattatccttattcctaaaaattagcaactactgtt ggacttaaaagaagggtttcaacttcctagtttagataaaacatatataaagtgggatttca ttagtggatgatgagtcattgcccggaaataaagcattttaacgtgatgaatcatttttca gaagtttttgaagtgccgatttggatattatttagaggaaatcctccgtatgtggatat aaagaaatggatgaaaaaacagcaaaaatacatatttggataaattatgaactcttttaac aggataaatttataattctacatttggaaaaagttattatttgaattctactctaaaaatcaga attttttcatttatcatgcctaattccatttttatttaattctactactctaaaaatcaga gaactaatcttaataatacatccatatataaatattgtaagaacttctgatgtgtttt aaagatgctaagtggaacctattattttaattctcaaaaaggatgatgaaggaaat aagactaagatacttataaaaaagatgatgtggatgaattccaataataattatttca gaacatttctttacacaagaagaatgggttgaaaaataattcaataataatttttttca gatgtatttactttttgattttattaaagaaaattgatggaaataatgaagaatttaattgat tatgtgttttttagttttaggtttaactccttatgataaataaaagttatgaagtatgagtgggat attattaaaaatagaaaaattccattcaaaaataaaaacttgacgatacttttaagaata ttgatgggtccgatattacaagataataatgtaaaaatgggcgaaaaagaataataaaa tatgggagctggttaggagcacctcgggaagaaaaattcttttaaaaatcctagaatttta ataagacagatttttctatagctcctaaggaatctcgaaaaaggatttttgcgcttac acagaagaagaattatataacgctcaaatgtcttttaatttagtattgaagaaggcttt gatgataaaaaatttattgaaatttttttaggaataaataattctaaaaatgatgactgg tattatgaagaaagattttatgataaaaaataaaaaagaattttgcaaaaattttgattgaa aatgctaaaaatcttccagtcattatttaattcaatttttttagatgagatagatctaatt gtggattcaataatagagtttaataaaagagtttttagtgaagaatgcatttcaaaaa tggcttaaaatagaatttgaaattgaaaaactctcaagaactagaactatttatgat ttaaactttgaagaattcttaaaagagataaaaaagaaaaagtagtcattagaccaaat caaatacaggactgtccgaattattttaatgaagtttaggaaaaatagaatatctgcaa agagagattaaagaagcagacgaaaaaaatttaacctactggtctatgaattatagtgtta	

FIG. 4A-12

GAMOLA ORF	Secuencia de nucleótidos/Descripción	SEQ ID NO:
2065	aatcatgaagaataagaaattatagaaaaatagctttaatgattaa atggacgatgaacattaatcctaattgaatacatcagaaacgcgccaacagagaaaaatggcttaagtcatttgaaggagtggaactt tataagaccaatccaaatttctagaaaaacaggcattcacccaaacaatgtcagcaagaaattaaaggatcttcgcaaacatgaactgg tctatgtcataaatccctgagtaccatgttcctaaactttacaggcttactgaaaagggtaaaaaatatgcttcaattctttag	141
	Sitio de integración del fago, locus de integración genómica	
2069	atgaataaaaaattatctttatccctccttttagtattattagtagctatttctgtctctgcagttgcagcagcagatgctgatgtcac atataaaacgatgctgcagatgtagacgatgttgacagcgaataaagtgtctctcttacagctagtgctgatgcacaagacatccaaa ctaagcttgataatgctaaacctggagacacaattgaattagaaaaacaagacatatgaogttgatatacaacatttaat	142

FIG. 4B-1

GAMOLA ORF	Secuencia de aminoácidos/Descripción	SEQ ID NO:
	Región genómica flanqueante, locus de integración genómica	
1980	vrmknkslllslslllitiisigsvvatdneeeinmdninnndnedianidhvdnvdnsninnpdtdiridnsnlreteldsnlnksn qiredelegsnaksnlksksllsttvdgsdenqmsnptiqsaisdanagdtiitgksyvhchfivnkphtiiiseigtmspcpsntk gsgahgifyispeasgtvlkgfnltntgyddygilirgaenveiinctintvsgdgirienatntkiadclikdsni	1
	integración cambio líticolisogénico	
1981	mktfrqllledpefnqyllqrpnltesslqsylnaatnfvrftgepfyktvhlrsqndriennliirfnpnqsrinimqfeyeylk grgctevsidsyvrmyrtllstlgiilpkspklddtpqdwyltkddikyldtanlqykavinfavtglrvdrmsltikdfmtate eyhgctevdfldsapdgmigwelfpqrtrkfrlpckvcntpessdllflsnervkyfewknekdgtdlkitknd	2
1983	mntlkiesckddyilkakansenelkvevpenwncdyvnavlweedicvlekgdermllipmcgelllegvqedeiykiyclpvkyq dqmvlaki	3
1984	mgmkvkieskaknmtrneklyfyklyddlynedrlilfstyfykffkddiilhyvlmnseneieeamkkideiqsegidiktffi pkkyckpkckvmdsygkicpdcgtliliedekkiqelqakdkayeeylekeyniylqnsyhnmiqgsytkirtkpkkteteivripagt vtsrgkfnssiqhypsetrytrqkvtrakykecrvlfdkemilnidgtatklydeveceleypeqevnelvtllhn	4
1985	mktqdliniindeespvlfnrevfendyvdpdiyykyrdeqlakmamyensipdnapiqlcggngatgkttlkqffkmlneafpnivt vyincqlfntentvygkiynklygkgsingksntmlfdkivarlkkenkiliigiddfsksrdglnkmlnlyflriheaeegiqici ftvsnkgeseslllpsrqsstgfyrlstpswrccttywtgapsvsila	5
	Replicación de ADN	
1988	mflekcdgkdtismlqekmnliletmeskgspfcisihcniipineaeewykngeigdqdfinfyddvnlieesfgfeyiyykseyptlh tqsinqiastypmnrtnqektplfirrerklyeitnifkshstneifismdsksklkheikydftrlkelneifknyleedcsiyilndn rafvmtlghfgefvdvfgsakesyvfdeiddekydkifirssyyfnipvdeldglaiklkqkriieegffegqfly	6
1989	miiflklkfrvimekllieidgvqyteaqirralaierdvrsnfvdmllgkikpselasrvsekga	7
1990	mcyvgntrtlvyhtedcfcnhllnenktilleekpvdmpcsfckpqfde	8
1991	mldmvaemvenirkgegdgysiyppfscivflgqkyseccckaeardqkfalvnlvgfredvkvdivprtneelfv	9
1992	mamladefeparlhkrtwaerhdveillaviclaisiamlllffalaeptvagvi	10
1993	mtkefedfmrntgllvflrwdtvalkyleshyderkyecayrllleaaidnlfdyqitfskktetereipdqlfekdkinkgflshigka ckkqsdgygrwkstrgirdayghysagnflfadiygcilhrisdccyrilnfseyeieeka	11
1995	mneiivtttennnvpdvdyaleewkayqrlltrelldetdyqthrgrkyktsawqkyarafnintqiidkeivkndkgivieaeytvra tlpngrfvesdgsdrrresgkremshsikataktratnralseligagdvdsadelpafdkvqshktnhvieaeaviespydknag fetadkiepvdvedpcknvwkticktikaeqkpcclkgvliqkartigtmtdeernrlieyiktlpkgevnldd	12

FIG. 4B-2

GAMOLA ORF	Secuencia de aminoácidos/Descripción	SEQ ID NO:
1996	mieciqeeqgftdwevpssssdikiyivsvdddegnlfcscpdyyrksrmphispesycckhirqvleednrlqml	13
1998	mdninkttkslakfeeffstvykdevmevlekypeedtlvdyenlemfddpdladlliekpdeviaasqkaiknidpmlmkdpkldikfk nvsncidfvnadskyigklisfeakvmeakepildiaavyecrgmslreipqtinssleplscpecggrsfrllqdeseflesqlli vssddtskslkvlllrdecdfldysmggevrilgiklsfssnygyeyflecnielindsedseydeygnrnspeyr	14
1999	mankirvnltdvpnlwqlakdklpcsrseffengqlkmflgiedeseiikdiqtkenenaldklchvrksqqlklesnksmekamas lnrmhkygkigenqirnlahvkvdfddllkceqdcnmcnifeaevpkhdsvm	15
2000	mtigslndfgksdgenmnpedfdcsvffemykalfeildvevgsfaelldvynvemdytlkrhalkkeilywfntdwkeelgkekpt ekdkekwiqrkigidsfvveqlevklkhirmyetalkhsfeaik	16
2001	mnvktvmndliglskefegveyeiesknsiyfysfpykymkegivilkysaiydlhtilkgmdgiivdilevednpgdekddllyvqviev ke	17
2002	misdeweeeyvkvneeleqvevrffskvdrlvfaqsyysssssfseaeailcdricdiltndlgnakyylggkdel	18
2003	mnyrewasqfelrmtedgrvcfktpcsyytfskedfeiiiremflnfes	19
2004	mdrikelnvogtckhshlipdingeiaavnicrigseavnkdggltycvdwtprkls	20
2005	mlskkeaiqmtldnekhypvkckycgkpfckshnrqmycsdsccrnalreqkaryqakrrlkikqkvliivdeykykglysgytsanghr knnfsheymaiakemkrigkr	21
2006	mylakfcpncgnkveendkfcicycgnklrvlipekkvkrssnsindektskyvevidglmykvfpsslpvkiylykvnygttsdeikn ilengnyynikinihyflqnkklyfrspknpnmffkfhndrfneemrkdnekeiishissyarvhrpsltkifnfnqektfilankhfeeni rktril	22
2007	miydkatvtslivailplmsmlgigeltqnyllaiivsgmialvwynekhnsdlvsgttkdcclcyggddeal	23
2008	mkhlyeiipyrtrtvtwittgflktvssamittgvilfnfnsitehpyfmewdeigivlgivsitiaiciyamidrwkerrkkeeeldtiety inrkaeeianmkvlrkleeelee	24
2010	mieistikitdikpaeynprimsqlehtklrnsmetfgvvdpiiinlknhiigghqryevllldksmednefikelhirlrgdvgwafp esdlevedddhekalnlalnniegewdplklepiltldkdvdfdieltgfsdieltelnennlvfaeefepdeseeedvlediydepv kemlqcpacdhvdvkrfrkrdvsqgd	25
2012	mdvdnllgydkvlelrsllleevtdk vipvwhknrgikdkfkmccqdyfnvsiisgwrnedvkdqfihfvrhahngcrihglgltrrkv ldrvpfdsvdssswlqtilyarlqgkqldskfaterqgdavlsyikwmktqeeyykkwrhyhd	26
2013	minerirpynlfnfndkkvifvtlfnvsnlislailaivfnlglfwglttdcgnllfplgylmadvitevygertarrvillglfanill ivattltvypypsywtggayayamfgftprivlagfiaylvqgvfnarlmvliikkwtnskylfmrtingstlgelcdscicssiaayyg ivpnsgilllilmqyvvkvvtwevmmqpltyksiaawarkdg	27
Encapsidación del ADN		
2014	mpepwerqrdengklepikafeyfteyltmdkprsmrvlcerlgkdkdyirqlhayssstwnwveraeydehiilkkrlrkekfydelv	28

FIG. 4B-3

GAMOLA ORF	Secuencia de aminoácidos/Descripción	SEQ ID NO:
	eselpnlrkrleynknmdietdmttkptsakahaydknskahsttllneillmigkpteiketsleadiesdnkidlenrvevditsde fmeseleymrknieek	
2016	mkdivnhygylspkptirsdskaknkfklnepyrqqlsagaggsimgygagllivddpiknvaeasvraqaklkdwgggtiksrvtq rrsnglppikiviahrlhldhgiiketepipandafirlrnggsidpntwvdfnlpaicdseddilgrkigevlweeqrdyewlma ekrsmgsylfnsiyggpverdgeifkrewfqdevnhkltclidpkdipqlrlywdfgasgdagdgtsailtsy	29
2017	MSKKQEmmrierlkhyayqtgliipifknhdlikkienkgrlkeyvriirreygldftsvklklkidnnkirskeelneeieekkt nfeiklsleelmneiesekkngrfknieekelikierriiefiniidnrdfkldyvtiqklvskigsneittkeelikem elrrivysdlnwldkweleskirhneittkeelikierriiefiniidnrdfkldyvtiqklvskigsneittkeelikem	30
2018	msknakadafvvttedgsydivdadvleryaiksesdetgskqlktgdweyddtlleplydplqlcelleintyhencvdrvardsagi gydivptgekekelnkpkltnfleniepninellyqmnrydratgygalelirkdksksepvnshissytlrtsdgrkrvkrvgtk tvwfiykgnydkegnldvhsetgefhpynslskeeranelwtmeyttksyglpkiygaipaiysdisrskyn	31
Módulo de cápside del fago		
2019	mmalkvrhskrqiqnikreyrrrrllleeqcsreiadffrrlerkihkvmdeweselglfhlknksdiidqsrqeydyilfkycckdsym kgreateerrfnrklenismkadvnitrlledlfpdptirynlnnkvfqasahmdrvdnrimenitqsyddglgideakdrltveyngl ksweaqriarteinsagndafdvyeigvghqwtadqervretpqadhrelhgkiykvgnsfnglqyppgdr	32
2021	mieipgndrteeinlpngqfvliylqendmmslpdgkyvcpfmiqlfteeggeligecieenphyntrfnytvfehldegieyr	33
2022	mlrvvertyyqqeektldcrrreagvntyslarqgavdyptyesynevreerikeakekygesyyhwrdrvetyfyllgrffsdleel ekylertvtykphreelkeamerldkrfeevinefwsyleeyeditdvleqlkegdctvhceflikefkkfvnvcricriikqneinhke f	34
2023	mdainvingnkilvdvlyrgtgnlidivigdalvydnptvvvkcyttldlafatteideivleneefellityedgefsillkshnlgel gyiewvi	35
2024	mvfvdlceneiadmvesfyrngdgsrrlltnravgeivehycsfevdgrittnlrdfllysvvlydtigeavddgvnleevfiedrnc ytgksqviliig	36
2025	mvnwlkaigdnfsvdyllalfssgdllilvaivlnsygvispenvrelvidiysyrkvdifwrhlrrprmsfedyvldnfeemetgelt reqvvefvsrqrkgltfcnelfiaivplkkgskddiveilwneyfvedykenwleqhenlgwndwkllkkeivenggdgdfqifrnlli dcvlmey	37
2026	mtmkvtfedengertvefgddvdvfliesdddnleiregdwelgdsgdddwdeedf	38
	mdefvetlfdtywkvnnengeymstldcggfyla kvapcvrnwsiviecnfcfhckefvyhengailleigmeisslylsqmeikdkliy midc	39
2028	mflegifeqdgenvreqviywrksnqvhnfvvnaqdgedncqphsvsreqleelrdlcravladndkaeellptprpgffgaiddyew yyydlqytvekidvylkddrylyfeycsww	40
2029	mqlvvegenmecpcdncmlreneepikvidwksnspfgngixihfwrerlqdvgyng	41

FIG. 4B-4

GAMOLA ORF	Secuencia de aminoácidos/Descripción	SEQ ID NO:
2031	melmtremegklksfysqdgkddaiavmkkfnpypglgtwyvlaekqngdylffgyvespitpefneygyfslselenlkikipiki ngitvsgyrierldlyfervirgdiagn	42
2032	mfgqkkefvkmyhvgdvvelvmdaqaappsgtrgeilfvddigqihvrwengsglaligyedrfkvverkge	43
2033	mdllylyddltarrevydvsglsfvvkykfssrkeaqdfalkygaelliee	44
2034	mierrrlgmkydiftlidelisrkliddgelsdeqvdfllqmeilveegtitedgaqdvmmngdy	45
2035	mmklslkelgeeiemi laeggltdqvdyllyletciadgsiteeqkreilicrdf	46
2036	mvregerlimyleidevkcenidrieiddvameivltdekvyerikrwlksneidydcreddryfanlieyviritw	47
2037	mvtvrdvgftieerffitaqeleysvgeehesvidraiallytklrrdfeftdeerelledafivsdq	48
2038	makenvidykierqndntwsylyvtergrgniaassfgelrqkvkrglpwndisnlfthkssdsnvrrsvkniddesvladvakss dsnvrlavrkisdnyvldivknasdyvrrreavrkindsdvlediaknnndenvrleavrnindesvlenisnksnadsdkvrieaikk indetiiaikkaknnndedvrieavrkinkdktviidfaknasdkvrreavrkindsavlaylknpswivrieavr	49
2039	mvkpcnfsprvskcdtnikwqcdckckfnhgafinaemekveqletiekiererterqferalkkeqfererteqfererke rlerekerekiekererkerlernrikieerererikrnerirrenernriksdkererekieerikrnerirkanernsirikrekr ererernvmtideyrsigystgskvswaiipillviciililmfygggm	50
2040	mtgqirqiaheylanyslvdknheffetrevigvpvesyitnepislkgdgtvneypkgtwiattritdeeemekalngeygtysit tvskkfadkqiqlprvmlkdkdpvgftislvrkpcvrgakfcmkediengdvsvseniddkieetkgfvqsikgfnkedkdedkn pededieldikavdevtkdfvntddfetfnelekalsdkfetlgaelfkslekdkaeaaaksgsr	51
2041	mgneatlnglvneqekavfksmrtmetgkavlneqlygflreatldntilrdadfkmlksfkxhlnrvgingrvtlntgydvngtdp eipaaadvdfganeldvkklkamceiedekednmtqaqfeqlmqmgerigedleywalfadsevarsddplntndgwlkkanhis srsiapsngmfdiedgpeamfdamikalpprfrknrrmlkfyvpfevedayrnilnrgtglgdsaqigfnalsykgipiehcstldde dgrgmngnrvcsn	52
2042	mrrrclnspehngmishiivllficfiglveailmalvdwedlaisvrksprklynvlkdelgipewnelsvierrrsmkkryavirdsfp elppweelsvidrrshkrlykklksvydyddpslesgppaavgpqkeipleaeayp	53
2043	mtwigtedviefgtvqpqtfrfekgdtssletllekwilgaegliisyncnydfndleeeippavvncvlrltanmvalaqarkdtpviqv kewnvqtvssnifsdnlkrdltpfvherksykgdeidffvitgddsdw	54
2044	mvklqidveelkpleprfkvkrvtvltanelqrnlkklspvdhgrlqgssvifqtgelertvksakyaifvndgtgylgplghkir pkngkflaftpnkgfkglvvpwptvrgqkqprfversmenterrvqefmirammemds	55
2045	mrfvntaslvpqtvkaylereicegglledvetlipsvnsdvpdpapaiwivqhppttrvsgsgnlsnkiamsvpfevcvveysddlee aeilgislasrvsgsslmknfnkvkvdssmpnrfhklefetlypvgevtvvgkseripatsiifnfyvvdwlkcnrryd	56
2046	mgirvvgmkeearygvaesapdfhqevskakaslntspntkssgsrmkkkaragvykptaniegevdllkrighylkaflndnyhftdggg nppvhefwggennklssftlwvtdifektivgslldnfkmevsdeymkftadvfykteesdeenielkykvkllldgdwalmfydvsvve idenappgivssfsfdgknlnvdkitiglsrgppqrkaaaqgrdisisfvstleretlelligkaeygevgtepeack	57

FIG. 4B-5

GAMOLA ORF	Secuencia de aminoácidos/Descripción	SEQ ID NO:
2047	mvvkkdilkqvkkiakvkiealdgdemylrplsqaeeinevdeieakamgifeetnetahgrrrqkpkvsvskgkinlelqqkqhqa ktkaiiflsldneknvgeeeawseteiegmphklfeelfnhvkrlsgielddedvdtfh	58
2049	mpssnmniivkaedmassvaqkvnsfrklngntidstfttslntkfnqeltsgtdldkvtqrlkqvgvngqssfnqltnaerrtle klsefdpvsaqvlqhlslrigitgqtfngqlsvseqslmmkstaqaaleevnsklrvvgigaaqaanmlnqmkldpsvsgnldraklkv semgysldstkgkilvltgtaiqtslgnkwdsiiktqvgttatnirtalgnaltskkskvqnlgnafsglgiissaig	59
2052	masvtkypsnvsgttgkfsvfslaniknnadgahavssvlikskkqspnrpstvsckgfsgfslpegaeptkitvtvyrhrknagsdys sknkhicniggtptisllgvsgfssksgcgtttmtthtkafsvhgklraqvnsanfgvklodyptnsntnygmrisyrvrtveyitsq ysvsvkhvsgtyededyvsvlsgisnknltsynptctltvpagftykgvtgaatgtvkvnrtfswnpqlqgragsrqislafepnvtf pegtdsy	60
2053	mgiaivvmneenflqfldpdlctinetieelglrtlefnykfqdyvedrdlfrignkiwisnsqslcdclvintpvensvyqenyfa ceieevlaelyyaplfstqelttsangftlrrtngqetvvdwnalnwfglyfnigvqecclgtyanritvngtmnrlnllrsieeqtg nrfvtryekdlldntihryldflnpvnvskwnknieydfiyeddegeyceaytsdgnpiseiyddieeeddivdfpp	61
Lisis		
2055	mvekitvqpqevrgygnvndekeledygyrcdvssesevikgveerifsvsgvpapalsianvtedtrrgcrakahisasfedgedgld dkaislksgdvlatittgsgenvdvlvlydsaqlyavfdgddyppavseaitvpakslwdeveldeeyevgdtaillsgvtgltiv deivdgeivtrrrqmeanvltlvtldgrrcstnangevflqvpniqngqrwvvaatsthlfnglidvpvhdsyl	62
2058	mvrfsrdmlqdgakrmfkwlrkgeglpnylimydmrknkeyklvpkeyaglyesrnfikwngrepyvlttsvarnplvmdyqntny ccptslslasqmlhykkesecakalgtsksgstspaqlianapklgfkliipikrdskevkkylkgfpvichwqvnqsrnckgdytgn fghygliwdmtsthyvvadpakgvnrkykfscldnankgyrnyyvvcpa	63
Conversión lisogénica		
2061	mkkhcfyflgdsfadieneamfcekhlvegnyldsiiragkaseiitvniaceleggdglissgqkkrlemlygkgyiisyydiykrlnhir kirnkavhghlsdienednanilhaylylicayfykeyrdtnfsaedtygpmidiaaskpetaasetnednenigefissplddyfkeyydd syllnellsklkdsskeavednlsfkeylhidrsiqedflkalnratsfnsshlmlcgsvgdgkshlianlkknpelfnqfaihyd atesfdpeknaidtlasvlepfdndnlnnsteklilainlgvlnflessyanedytklklieeanifesnevsdnygdkvsvftfs dynmfelnddensnytskkyisslfnkitqkedtnpfyaylkkdshfinpiinyemlmdeevqktiidylikifkyrkiistrdl lnfiyeiivppeflksedldnindmdyslpnllfgypersdllklcneldptlhrnesldkfiidlninddtekilnryfdtrfnfl eeeygeylvdfrefnnsekekvtnilirfavfygksiiiknfdkvylnylkyayntqshdkykylftevkdafnwksyknktici dtldsfcvyknllklkpsvdkfeksldglflgnrfrktidikiyfsvesnkkkplnvdflslyqymklyngfkpnqsdckdldlildefin nlldedtdddlyvisletyeeelfesndfgtfefkrg	64
2062	mdfsenyinillkqmtcdvnrkrlhinqnspplpfbktntpkkanfengfdiilgelsrillnktiekfnkldnivsnlindnieiedg tkeyitklllneylfdekndlkishpnlylyiplsnkssngeqevalfrdicknngnlinffesydsnhiilnllkntpnllhkit	65

FIG. 4B-6

GAMOLA ORF	Secuencia de aminoácidos/Descripción	SEQ ID NO:
2063	etkyvihfeeianlfnedinyaillykffmenignifayyyffvisqlilkiskgfdndnefeklyylldwesasknrkslnsysllkh hskplyakmavidqintllgttnlllekdiseyfnldinsknflhflkwwsdyrvrnfddkelpdnllleilfeskneklgvd gavqsyalnlediaakyllkrrgsygyvlninrdmllvltalcvkdkkiklnqlfiefekrgvyfdkyskeevnfltklnldkksd sgdaqyvkpvl mlnqfydylsnkllnyfddtkilsgckffisfdeaddqimsfynslrsiaetnscsefiyvhtisgkeyntysiningvkfviseslti nvdfvltrnqvtsqegvkwkdtallvicneaisdigkmrnlqkegmplnvksisnledeindsqilnysdkqiafkfslniqeeelfq ttlwdyetilsinkgfvdsdedlrelnlfkddqlnqspqkmlkrlkenydtfnevnkfsqygdkkeqlknmftdsgvilskddwyka ewkmvkskddfinqgnplnynenlekitenglnyweipnsftktgkrkniivfnphshesvlkfsdqilsnslntnskkftiar gksliivnftldssepifktiikykhkneniseftnfvlnfepeifnsiksrsfsvnvksqilvntndedsdivfgtgskeiekliken geklylyddesliiseqspawndgklsfkykdnyapflikekskktlpvnsyviwnlkrnmenfifngvkavqdvnsfylvveefke flkmereiiikqdifyakrnidgslekiesvsnleataymdifnyyktfddspednlpslvlynddlkelykkfiftnkeiseiens ilsdfkykknllklgrietdkimysplspniayqlvskqcgndesvnilerlvnnlipycsddgkelfrpiyqeeahewliye kseevsigttvnfisinvtetklnqfkhfnylfsfnsspi kinlinikddkevkvgnfirsrldpdkttkkipveiniyndaeks sfdnldfcqseiqlleefgikklksdifpidi ihmiqnisiyykhpfkkeeyeahlsfykvkshnniandmmdkietglslngllss vtsttkhseyrtgftgnnlmsnpliktvlnlnelvensknfgntysknksvittveleednieelydkshwtvfiptfgieyfds sdsnlhiihysdqysssskdyttitvtnkstqveeiiirdflgskyykvtddeelyvikmfnsingewllrvinsnghydraklsisaik yclsildhkddivwipvmeeilriagnvklcknkgifdsklkgnhsddllfigvknednrievifypievkiglnnastikkgsq dntykllktqlninvensefrnkfrnffiqlllneqklvtnhidekgldrieefkaellndeydilygleeyigkgslnvsfkkes hhisiymdvdkqvlelpedfayyglatpreihdeiqsdntdilaetllshvdiseiraknndicdsnedmsiddfdnlsefedsfie eeseiseepdeeltatsssdnesienigespskisnvralligtqkgnhkvwywefghpslanrhmligksggktyfigrmlkelsi qgipsiidiytdgfkpsqlpnfkdsldgklsqyfvvkenfpinpfkrntimidkdiieednstiasrfksiinsvyglgiqqsntly qtvldcldkyddnfdlnlkeeilkdesnaqtvlknlnelldknpfastdfdwsvldnkdkgviiqltalskdiqitiitefilwdlw nykltngsednpiivldeahnlfdsndspcskilkegrkfgwsgwfatqsvkgsmkideiaklenadekiyfhtpdvstiaaklskdn edkkiyekelsqltkgycivqgsaidssgnlyqnpvkvieeisfden meskdrientifndkrmdayidkyfeeftlssequeaalniwidklnndqltseknynhnffeiiledllgykrsvdkheenigdeghev efvlekdgdyviielkgtykdltkrrpgqspveqatnyasakketewatvsnnydefrnfntardnyisfkfrqlkdleifkklfl1 vfskfsliidedipkllnetkvierelenefyqlysdtrlmiikeleysseidnriekalsqilnrfiflcaedlameeettadv lltplkhrnligntmwnrlnelifanggnkhrripafngglfeddlnlkiideidrsfvenwnlkedfedkyediakligvykdtl npifinlliiistydfdselvnlighifensisdieelkndnqeqrkdqvytytpeyitdyicrntiipylsisgkastvhellyeyes snsltdvldsklnikvldpacgsgmlnksvdilfeihealhasakyagdsldrffdslekrkeiisnniyygvdlnesveitklsfl1 klattvlgkegqlpsldkhikcgdsldvdesiagnkafnwyesfsevggfdiivgnpppyvdikemdektaiyifdnyetsfnrin lystfveksyylknegifsfimpsilfnstyskirelilnntsilnrvtsddvfkakvepiilifkkydegntkilkikddmd eipinnyshefftqerwfennsiinifsdtdftfdllkkidgnnerlidycdflsgltptydkykgmsediiknrkfhskiklddtfkell dgsdityrvnkwekeyikygdwlgapreekffnprilirqilsiapkeskrifaayteeelnyaqiafnlvkdegddknllykfl	66
2064		67

FIG. 4B-7

GAMOLA ORF	Secuencia de aminoácidos/Descripción	SEQ ID NO:
2065	giinskmmtwyeeerfmdknknknfakiliienaknlpviinsnfldeivsnvdsiieInkefysvrnafgtwlkiefeleklskklenyy dlnfeefllkeiikkkkvirpnqldlselfneslgkieylqreikeadekinllvyelyglnehheieiensfnd mddetlilileiyirnaptremvlsfegvdfirpiqisrktgihpnnvskklkdlrehelvyvinpeyhvpklyrltekgnmlqfl	68
	Sitio de integración del fago, locus de integración genómica	
2069	mkkkiilsllllvllvaisvsavaaadadvtyindaadvddvadekvapltasadaqdiqtkldnâkpgdtielenktydvdttnvntkq vtikgqdtvikasgasggsgalfianeagtafegitfintdghknygeqvsyayiqalaiengtvdnckfidwssgvvykgasfcsitn syfngsseqvtnggkkeygtkainlmgshditvtgctfegqvlдайsiaasnsgnnimtdntfidncyaiyfggastq	69

FIG. 5

	(101)	110	120	130	140	150
PeiP (98)	SRYNRFKEVNGREPRVVEITYSG--CGPSVSLRTFKDMCKRYNQFL	EEENRR				
PeiW (83)	RRYEDFVPTNGREPNLITYLEQG--KSDHVSLSCTFKDMLRRYKDFVR	INGR				
MRJ ORF2058 (74)	-----VARNPLVMDYQNTNYTCCTSLSLASQMLYHYKSES-----					
Consenso (101)	RY F INGREP VIFI NG	VSL TFKDMLKRYK FL N R				
	Sección 4					
	(151)	160	170	180	190	200
PeiP (146)	EPRIYVYTPPEPPVPEVRERVRVLCFFK-TATQLYTLVSRRCYKRYFYYN					
PeiW (131)	EPNYTSTQPPSLKCHWTTKVTEKTCFPH-DATSLYERVRKTCYKRYYYN					
MRJ ORF2058 (110)	-----EAKALCTSKSGTSPAQLTANAPKLGFKII					
Consenso (151)	EP I I P	EM K LGTFK SATSLY LVAK CKYKRYFYYN				
	Sección 5					
	(201)	210	220	230	240	250
PeiP (195)	DQTPNREALKKMMTDCINCTDACQLFKPVTEGLCYSVRIEHVKVRCNDNK					
PeiW (180)	DQVPNHVAVMMTTSCINCTDACQLFSKVEEMGYEVKI EHVVRVKCNDCK					
MRJ ORF2058 (141)	PIKRD SKRVKKYLRKCGFPVICHWQVNSRNCRGDYTGNF-----					
Consenso (201)	DQ PN AVKMLT GINCTDACQLF	VIE LGYSVKIEHVVKVRCND K				
	Sección 6					
	(251)	260	270	280	290	300
PeiP (245)	WYCHYFLRVAGCKELASVSLPSERWTVWDYVSAATKTCRPLCAPCCS	RGIOH				
PeiW (230)	WYCHYLLRVGCFELKDC-----TWWDYVSAATKTCRPLGVPCTACFOH					
MRJ ORF2058 (180)	--CHYCLIWNTSTHYV-----VADPAKGVNRKYFSCLDNANKGYR					
Consenso (251)	WYCHY LRVAG EL V	TWWDYVSAATKTCRPLG	PCCS G QH			
	Sección 7					
	(301)	312				
PeiP (295)	LGWGI VSPKHD--					
PeiW (273)	LGWGI VGPVYDK					
MRJ ORF2058 (220)	QNYMVC PA---					
Consenso (301)	LGWGI V P HD					

SEQ ID NO:70
 SEQ ID NO:71
 SEQ ID NO:72
 SEQ ID NO:73

FIG. 6

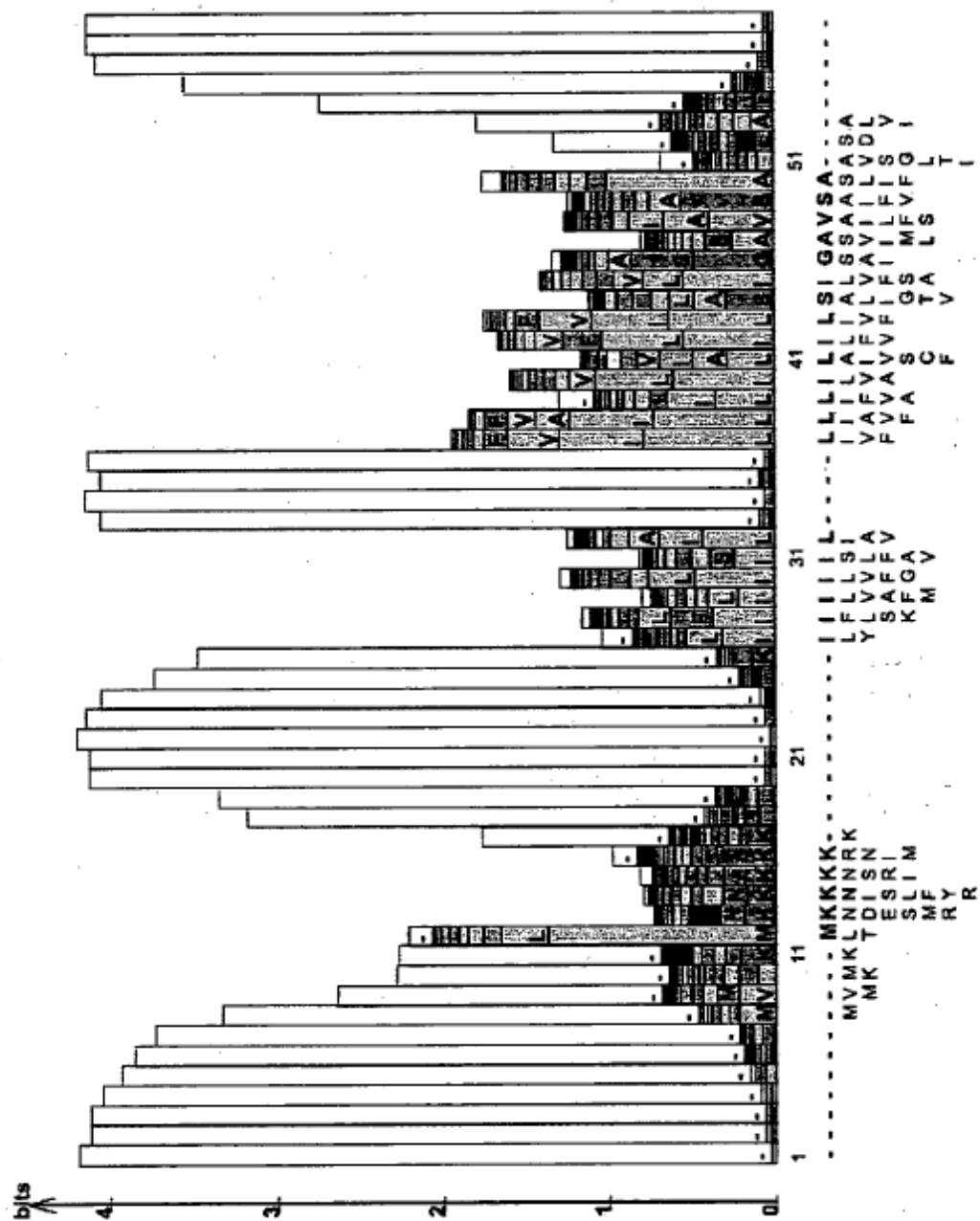


Figura 7. Efecto de la enzima lítica ORF 2058 en células de
M. ruminantium en reposo

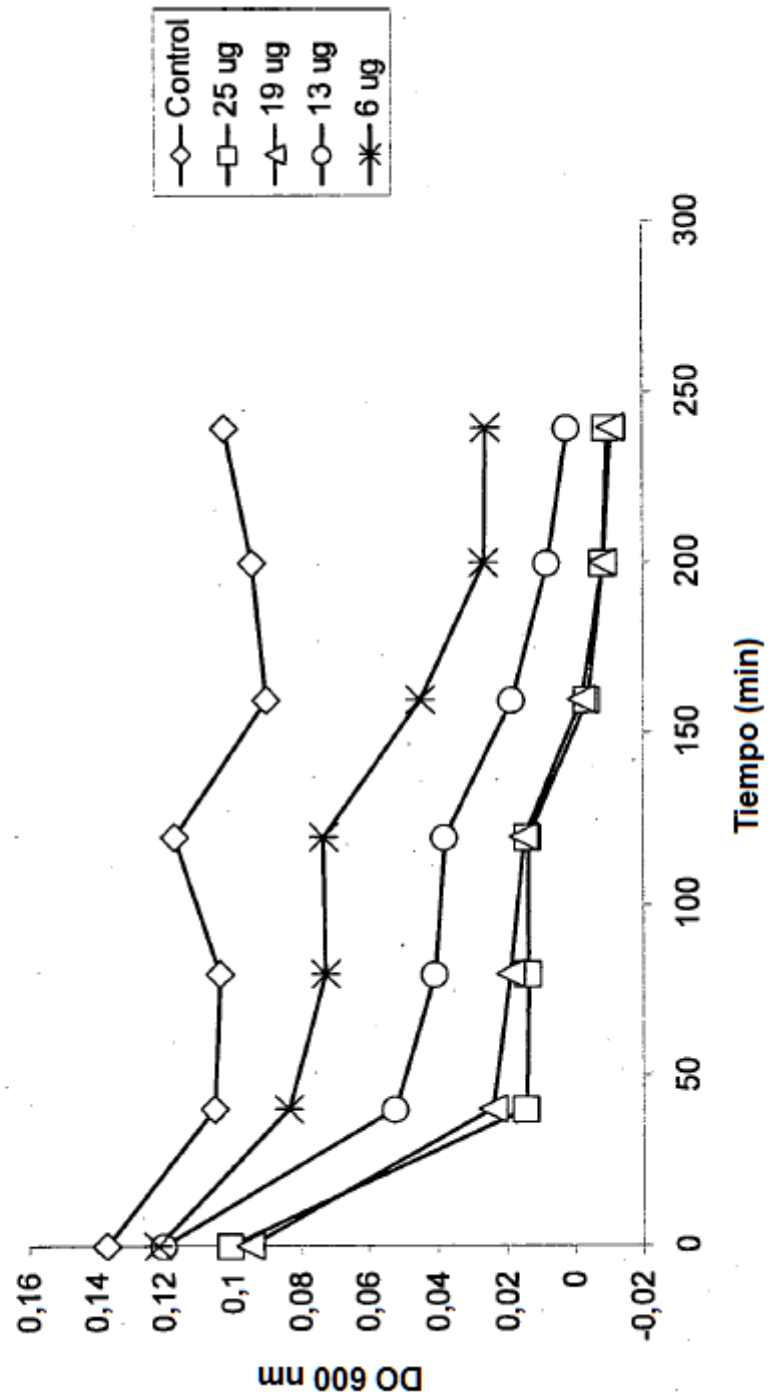


Figura 8. Efecto de la enzima lítica ORF 2058 en el crecimiento y formación de metano de *M. ruminantium*

